STEM CELLS ARE HIGHLY ENRICHED SOURCES OF AUTOANTIGENS TARGETED IN SCLERODERMA – IMPLICATIONS FOR PATHOGENESIS

by

Tricia Ruth Cottrell

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

March, 2014

Abstract

Myositis, scleroderma, and other systemic autoimmune diseases are united by many features, including (*i*) the presence of a self amplifying cycle of tissue destruction and immune system activation with a failure of homeostatic tissue repair; (*ii*) an increased risk of cancer; and (*iii*) a strong association of particular clinical phenotypes with disease specific autoantibodies, which are used clinically for diagnosis and prognosis of disease. Interestingly, most autoantigens identified to date are ubiquitously expressed proteins, despite their associations with highly specific patterns of tissue damage. Finally, there is a significant subset of patients with characteristic features of systemic autoimmunity that lack any known autoantibodies. While most autoantigen identification studies to date have used tractable, immortalized cancer cell lines, we proposed that these cell types likely express only a small subset of proteins targeted in systemic autoimmune disease and the use of appropriately perturbed antigen sources that mimic changes in the target tissues are key to the discovery of novel autoantigens.

Studies in autoimmune myositis focused on key characteristics of skeletal muscle in this disease, including the presence of a type I interferon signature and increased antigen expression by regenerating muscle cells, to identify two novel autoantigens, IFIT3 and MYL4. These studies demonstrated a systematic triangulation approach to improve the efficiency of autoantigen identification, featuring: (*i*) antigen sources expressing pathways shown to be active in the diseased tissue *in vivo*; (*ii*) confirmation of spot selection by post-pluck immunoblotting prior to LC/MS/MS peptide sequencing; and (*iii*) the use of data from additional methods to filter the proteomic data, prior to rigorous experimental confirmation of antigen identity.

This approach was then applied to the study of scleroderma. Based on the disease model that an antitumor immune response may be misdirected against progenitor cells involved in tissue repair, we sought to prove that differentiating progenitor cells are a target of the immune response in scleroderma. Our studies demonstrate that the expression of known scleroderma autoantigens is altered during stem cell differentiation. Moreover, immunoblot screening of lysates from undifferentiated and differentiating stem cells with scleroderma patient sera revealed that roughly one third of randomly selected patients and 60% of ANA negative patients have autoantibodies targeting progenitor antigens, which are not expressed in HeLa cells and not targeted by autoantibodies in healthy controls. Finally, two of these novel autoantigens were identified as keratin 8 and fetuin A.

These studies outline a systematic approach to autoantigen identification that accommodates the complexity of protein dynamics in perturbed cells and limits the investment of time and financial resources necessary for the identification of any particular target. These are the first studies to implicate differentiating progenitor cells as a target of the autoimmune response in scleroderma. Finally, this work demonstrates the importance of screening antigen sources that approximate changes in the target tissue in order to identify novel autoantigens that have potential clinical utility as well as the possibility of providing mechanistic insights into these potentially devastating and poorly understood diseases.

Thesis Advisor: Antony Rosen, M.D.

Thesis Reader: Deborah Andrew, Ph.D.

Acknowledgements

I am incredibly grateful to Dr. Antony Rosen for his mentorship and support throughout my graduate training. The depth of his wisdom in scientific, professional, and personal matters cannot be overstated and I have no doubt that I will continue to benefit from his influence throughout my career. I am specifically grateful for his emphasis on perseverance, rigorous experimental design, and the importance of addressing the most difficult scientific questions in order to improve our understanding of complex human diseases.

I would also like to acknowledge the tremendous contributions of Dr. Livia Casciola-Rosen to my training and completion of this thesis project. She is a wonderful mentor, providing both professional and personal guidance. Her willingness to share her expertise and provide moral support was essential to my success in the lab. Dr. Casciola-Rosen also had a crucial role in the partial publication of this work and continues to play a pivotal role in guiding ongoing experiments and manuscript preparations.

Several other members of the Johns Hopkins Division of Rheumatology have provided incredible support and mentorship throughout my graduate training. Dr. Francesco Boin's expert guidance through a clinical research project resulted in a publication and the acquisition of invaluable skills in biostatistics and data analysis. Dr. Fredrick Wigley's leadership of The Johns Hopkins Scleroderma Center and enthusiastic support of research was instrumental for the execution of these studies and provides an exemplary model of collaboration between clinicians and scientists striving to improve patient care. Dr. Erika Darrah Gattens is a friend and role model whose dedication and love of science is inspiring. I have benefited tremendously from her selfless efforts to

ensure the success of those around her and her passion, conviction, and scientific rigor make her an invaluable colleague and mentor.

I would like to thank all of the past and present members of the Rosen Lab for sharing their expertise. I owe particular gratitude to Dr. Zsuzsanna McMahan, who has enthusiastically pursued the continuation of this work; Dr. John Hall, who made significant contributions to the partial publication of this work; and Dr. Kimberly Maurer, who has been an amazing friend and colleague throughout our training. Dr. Andrew Mammen was also particularly helpful with the fluorescence microscopy and thoughtful career discussions. Tonie Hines, Janelle Montagne, David Hines, Brandon Boring, and Lesly-Ann Samedy were essential in keeping the lab running smoothly and I am grateful for their willingness to provide assistance under any circumstance. Other members of the lab who were helpful in creating a wonderful atmosphere for scientific discussions include: Dr. Katherine Pak, Dr. Laura Guitierrez, Dr. Jennifer Mammen, Dr. Felipe Andrade, Andrea Fava, Dr. Violeta Romero-Morales, Dr. Sun-Young Oh, Dr. Thomas Grader-Beck, and Dr. Tomeka Suber. I would also like to acknowledge Jolene Patey and Krystal Burgess, who have been instrumental in helping me navigate the administrative aspects of both scientific investigation and graduate training.

This success of this project depended on the efforts of numerous collaborators and colleagues within the Johns Hopkins community. These studies would not have been possible without the patients, staff, and physicians of The Johns Hopkins Scleroderma Center, especially Dr. Fredrick Wigley, Adrianne Woods, and Tanya Moore who were instrumental in providing the clinical information and patients samples used in these studies. I would like to thank Dr. Francesco Boin, Dr. Robert Wise, and Dr. Fredrick

Wigley for their contributions to the clinical study of skin disease and lung function in scleroderma patients, as well as Dr. Sharon Ghazarian of the Biostatistics, Epidemiology and Data Management (BEAD) Core at Bayview. I would like to acknowledge the Johns Hopkins Stem Cell Core Facility, led by Dr. Hyesoo Kim, for the culture of the human embryonic stem cells used in these studies. I would also like to thank Dr. Elias Zambidis and Dr. Tea Soon Park for providing endothelial progenitor cells as well as Dr. Michael Shamblott, Dr. John Zhan, and James Pendleton for the human embryonic stem cells used in the early stages of this project. The proteomic sequencing experiments and analysis were performed by The Johns Hopkins Mass Spectrometry and Proteomics Facility, including Dr. Robert Cole, Dr. David Colquhoun, Dr. Tatiana Boronina, and Lauren Hitt. The microarray studies were made possible by The Johns Hopkins Microarray Core Facility. I am very grateful for the assistance of Dr. Ann Hubbard and Lydia Nyasae in developing the embedding protocol for embryoid bodies as well as Dr. Allen Myers and Holly Rohde of the Bayview Histology Core and Jennifer Sipes of the Mouse IHC Core facility for tissue sectioning services.

I would like to thank all of the members of my thesis committee for their valuable guidance throughout the years, including Dr. Deborah Andrew, Dr. Charles Lowenstein, Dr. Michael Shamblott, and Dr. Hy Levitsky. I would also like to thank Dr. Deborah Andrew for her thoughtful suggestions for revision of my thesis manuscript. I am also indebted to the Johns Hopkins University Medical Scientist Training/MD-PhD Program, including the directors: Dr. Robert Siliciano and Dr. Andrea Cox, as well as the wonderful program administrators: Sharon Welling, Bernadine Harper, and Martha Buntin. I would also like to thank all of the members of the Graduate Program in

Cellular and Molecular Medicine, including the director, Dr. Rajini Rao, and very helpful administrators, Colleen Graham and Leslie Lichter. These members of the Johns Hopkins community have facilitated a wonderfully supportive learning environment throughout my training.

I would like to thank all of the friends I have made throughout my training, including those in the MD-PhD Program as well as my colleagues in the Johns Hopkins School of Medicine, the Graduate Program in Cellular and Molecular Medicine, and the Rosen lab. It has been a privilege to train with such a wonderful group of intelligent, caring, and hard working people who entertain and inspire me on a regular basis. These friendships have been instrumental to my personal and professional satisfaction over the years and I will always be grateful for the opportunity to have known each of you.

Finally, I would like to acknowledge my wonderful family for all of their love and support over the years. My mother, Patricia Gage, always emphasized the importance of education, always believed that I could accomplish my goals, and made many personal sacrifices to create opportunities in my life. My grandparents, Rhoda and Jim Ackley, were a constant and positive influence throughout my life and have greatly shaped the person that I have become. I thank my sister, Alyssa Cottrell, for always accepting, challenging, and inspiring me. I would also like to thank all of the other members of my family for creating an incredible foundation for my success and happiness. Finally, I would like to thank my husband, Steven Handy, for the love, support, and joy that you have brought into my life.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2D	two-dimensional
ACA	anti-centromere antibodies
AEP	asparagine endopeptidase
ANA	antinuclear antibodies
bFGF	basic fibroblast growth factor
BSA	bovine serum albumin
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
DM	dermatomyositis
DMEM	dulbecco's modified eagle medium
DTT	dithiothreitol
EB	human embryoid bodies
ES	undifferentiated human embryonic stem cells
EDTA	ethylenediaminetetraacetic acid
FDR	false discovery rate
FETA	fetuin A
HED	hydroxyethyl disulfide
HEK293	human embryonic kidney derived cell line
HeLa	human cervical epithelial cancer derived cell line
Hep2	human laryngeal carcinoma derived cell line
HI FCS	heat inactivated fetal calf serum
HMGCR	3-hydroxy-3-methylglutaryl-coenzyme A reductase
HRP	horseradish peroxidase

HRS	histidyl tRNA synthetase
HSC70	heat shock cognate 71 kDa protein
HSG	human salivary gland cell line
IEF	isoelectric focusing
IFN	interferon
IFIT3	interferon-induced protein with tetratricopeptide repeats 3
IPG	immobilized pH gradient
IVTT	*in vitro* transcription and translation
KBM-2	keratinocyte basal medium-2
KRT8	keratin 8
LC/MS/MS	liquid chromatography tandem mass spectrometry
MDA5	melanoma differentiation-associated protein 5
MEF	murine embryonic fibroblasts
MEM	minimal essential media
MHC	major histocompatibility complex
MS/MS	tandem mass spectrometry
MYL4	myosin light chain 4
NCAD	N-cadherin
NP40	nonidet P-40
PBS	phosphate buffered saline
PIs	protease inhibitors
PM	polymyositis
RA	rheumatoid arthritis

RPC1	RNA polymerase III subunit C1
RT	room temperature
SEREX	serological analysis of recombinant expression cDNA libraries
SERPA	serological proteome analysis
SLE	systemic lupus erythematosus
SSc	scleroderma
TNNI	troponin I, skeletal slow twitch
TRP	tyrosinase-related proteins
U2OS	human osteosarcoma derived cell line
VEGF	vascular endothelial growth factor

Chapter 1: General Introduction

Systemic autoimmune diseases

Human autoimmune diseases occur frequently (affecting in aggregate more than 5% of the population worldwide), and impose a significant burden of morbidity and mortality on the human population *(1)*. Autoimmune diseases are defined as diseases in which immune responses to specific self-antigens contribute to the ongoing tissue damage. Both the specificity of the immune response and its role in tissue damage are central components of the definition. Unlike tissue-specific autoimmune diseases, where unique tissue-specific antigens are targeted, systemic autoimmunity is characterized by multiple affected tissues and adaptive immune responses to a variety of apparently ubiquitously expressed autoantigens *(2)*.

Although the definition appears relatively simple in concept, the complexity of this spectrum of disorders is enormous, and has greatly challenged elucidation of simple shared mechanisms. This complexity affects almost every domain, including genetics, phenotypic expression, and kinetics. In the latter case, there is frequently a prolonged period (weeks to months) between initial onset of symptoms and development of the diagnostic phenotype, and disease may vary in expression in the same individual over time.

In spite of this enormous complexity, there is a striking association of the clinical phenotype with the targets of the autoimmune response. This association is, in fact, so strong that autoantibodies have been used for diagnosis and prognosis in the human autoimmune diseases *(2)*. For example, autoantibodies to the Sm splicing

ribonucleoprotein complex are diagnostic of systemic lupus erythematosus (SLE), and autoantibodies recognizing topoisomerase 1 are found in patients with the diffuse form of scleroderma. While not thought to be directly pathogenic in most cases, the presence of high titer, class-switched autoantibodies requires suprathreshold concentrations of molecular epitopes not previously tolerized and encountered in a pro-immune context. Thus, the association of specific autoantibodies with distinct clinical phenotypes provides critical clues to understanding the mechanisms underlying the initiation and propagation of autoimmune diseases.

Distinct phases in the development of autoimmunity: initiation and propagation

By studying the development of autoantibodies over time in patients who subsequently manifest an autoimmune disease, significant data have demonstrated that the onset of an autoimmune response and the development of clinical symptoms are generally separated in time. For example, rheumatoid arthritis (RA)-specific autoantibodies recognizing citrullinated proteins frequently precede the development of RA *(3)*. A landmark study in SLE expanded our understanding of this phenomenon. Harley and colleagues analyzed sera collected from patients in the US military who subsequently developed SLE *(4)*. Strikingly, patient autoantibodies could be divided into two groups: (*i*) those which precede the diagnosis of SLE by several years – these included antinuclear and antiphospholipid antibodies; and (*ii*) those which occurred around the time of onset of symptoms – these included anti-Sm, anti-RNP and to a lesser extent anti-DNA. The observation that one group of autoantibodies precedes symptoms in SLE, and that another group appears coincident with the phenotype strongly suggests

that the groups mark distinct events in the development of autoimmune disease. Members of the first group are likely markers of disease initiation; members of the second group likely markers of disease propagation. The antigens targeted by the immune system in this latter phase (i.e. associated with clinical disease) are more likely to themselves have some function in disease-propagation, possibly through their possession of pro-inflammatory or adjuvant functions *(5)*.

Initiation phase: protein dynamics and determinants of antigen selection

Initiation of an adaptive immune response requires presentation to T cells of suprathreshold concentrations of molecules with structure not previously tolerized by the host. One of the more persuasive models proposed to explain the persistence of potentially autoreactive T cells within the repertoire of the host is that of immunodominance of T cell epitopes *(6, 7)*. Studies by Sercarz and colleagues*(7)* have stressed that while antigens contain numerous potential determinants that could be presented on major histocompatibility complex (MHC) class II during antigen processing, not all determinants in a particular molecule are equally likely to be efficiently presented. Those determinants that are most efficiently presented are termed 'dominant'; those that are not loaded onto MHC class II to a significant degree are termed 'cryptic'. For self-antigens, it is likely that a constant set of dominant determinants are generated during antigen processing under most circumstances, with similar outcomes in the thymus and periphery. Antigens processed by the 'standard' pathway are therefore fully tolerized, with the T cell repertoire purged of reactivity to the dominant self. However, the balance of dominant and cryptic epitopes presented during

antigen processing is influenced significantly by changes of protein structure, which occur during various relevant physiological states *(8)*. Currently, there is evidence that post-translational modifications, proteolytic cleavage, and genetic changes altering protein structure are mechanisms that may alter antigen processing to reveal potentially cryptic epitopes in the initiation of systemic autoimmunity.

Post-translational modification of autoantigen structure

Autoantigens undergo a variety of post-translational modifications, including phosphorylation, proteolytic cleavage, ubiquitination, transglutamination, citrullination, and isoaspartyl modification *(9)*. In several cases, autoantibodies recognize exclusively the modified form of the antigen (e.g. RNA polymerase-II large subunit, SR proteins, citrullinated vimentin and other RA autoantigens), indicating that the modified forms of the molecules are important in driving the immune response. In addition, Doyle and Mamula *(9)* have demonstrated that post-translational modification of autoantigen structure may be more broadly relevant than can be appreciated by studying autoantibody specificity alone. They showed that whereas mouse immunization with a murine cytochrome c peptide (amino acids 90-104) resulted in no T or B cell response, immunization with the isoaspartyl form of this peptide resulted in strong T and B cell responses. Although the autoantibodies elicited recognized both the modified and the native form of the antigen, the T cells only recognized the isoaspartyl form. Similar observations have also been made for several SLE autoantigens *(10)*. The difficulty detecting and quantifying antigen-specific T cells in various autoimmune diseases may reflect their preferential recognition of subtly modified forms of autoantigen.

Novel antigen cleavage during cell damage, cell death, or inflammation

Recent studies have provided evidence that single proteolytic events early in antigen processing may play critical roles in defining the epitopes generated. For example, Watts and colleagues (reviewed in reference *(11)*) have conclusively demonstrated that early cleavage by asparagine endopeptidase (AEP) determines subsequent proteolytic events. Modifications of the antigen that affect this early cleavage dramatically change the epitopes loaded onto MHC class II *(11)*.

Inflammatory microenvironments can create significant potential to load distinct epitopes because unique proteolytic activities are present. Activated inflammatory cells constitute a major source of proteases, including various cytotoxic lymphocyte granule proteases (granzymes), as well as numerous neutrophil and monocyte granule proteases. It is of interest that numerous autoantigens targeted in systemic autoimmune diseases are substrates for these inflammatory proteases, and that unique autoantigen fragments are generated through the activity of granzyme B and potentially other similar proteases *(11)*. Such autoantigen forms are not generated during other forms of cell damage or death, and similar activity is not observed against non-autoantigens. Thus, novel proteolytic cleavage of intracellular autoantigens during activity of cytotoxic immune effector pathways may provide a source of cryptic epitopes not generated during homeostatic 'tolerance-inducing' tissue turnover.

Autoantigen alteration due to mutation, truncation, or splicing

Since the final epitopes generated and loaded onto MHC class II can be profoundly influenced by single, early cleavage events during antigen processing, relatively minor but critically placed changes in the primary structure of autoantigens may have the capacity to influence peptide selection. A study of the melanoma and vitiligo-associated autoantigens, tyrosinase-related proteins (TRP) 1 and 2, has strikingly demonstrated that this mechanism may play an important role in generating immune responses to self and tumor antigens *(12)*. In this study, Houghton and colleagues generated cDNA libraries encoding large numbers of random mutations in syngeneic TRP proteins and used a DNA immunization approach into black mice to test the immunogenicity of the altered proteins encoded by the pools of mutated cDNAs. Immunization with non-mutated proteins induced no detectable immune responses, but the mutated cDNA pools elicited both autoimmune depigmentation and the ability to reject melanoma tumors. Additional analysis showed that autoimmunity resulted from mutations that altered autoantigen cell biology, particularly degradation rates and pathways. Mutations also created new helper T cell epitopes, and induced recognition of non-mutated but previously cryptic epitopes. Interestingly, mutations themselves did not form part of CD8 epitopes that drive the anti-self and anti-tumor immune responses. Mutated molecules that were immunogenic were frequently truncated, leading the authors to propose that inappropriately truncated self-proteins may provoke autoimmunity when present in a pro-inflammatory environment.

Propagation phase: chronic inflammation driven by target tissue antigen expression

The striking association of specific autoantibody responses with distinct phenotypes suggests that autoantigen expression or form in specific target tissues may play an important role in both focusing the immune response and generating tissue damage. Recent studies on human autoimmune myopathies have begun to provide important insights into this problem. Myositis-specific autoantigens are expressed at very low levels in control muscle, but at high levels in myositis tissue, where antigen expression is at highest levels in regenerating muscle cells.*(13)* Interestingly, histidyl tRNA synthetase (HRS) expression is also found at high levels in lung and anti-HRS antibodies are associated with autoimmune myopathies with interstitial lung disease *(14)*. Recently, 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGCR) was identified as an autoantigen in patients with statin-associated autoimmune myopathy *(15)*. In addition to the induction of HMGCR protein expression by statin treatment of cultured cells, this group also demonstrated increased expression of this protein in regenerating muscle fibers of anti-HMGCR positive patients. These data suggest that enhanced autoantigen expression in the target tissue may be a feature of disease propagation, and that antigen expression during tissue repair may provide an ongoing antigen source to sustain and amplify tissue damage.

Systemic autoimmunity and cancer

There are significant data linking autoimmunity and cancer. For example, cancer is present in ~20% of individuals with dermatomyositis, with a striking temporal clustering of cancer diagnosis around the time of diagnosis of the myositis *(16, 17)*. Similarly, there is evidence for an association of SLE with cancer, particularly

lymphoma, again clustered within the first two years of SLE diagnosis *(18, 19)*. Sjögren's syndrome is also strongly associated with cancer—the relative risk for lymphoma in Sjögren's patients is roughly 40 times that of the general population *(20)*. Rheumatoid arthritis patients develop lymphoma at double the rate seen in the general population, a risk correlated with disease severity *(21)*, but not the use of biological therapies *(22)*. Finally, a comprehensive review of the literature confirms an increased risk of malignancy in scleroderma. In this disease, lung cancer is the most common malignancy, but the diagnoses of breast cancer and scleroderma show significant temporal association *(23, 24)*. These associations, both with timing of diagnosis as well as preferentially with specific tumor types, are strongly indicative of a non-random clustering of autoimmune processes and cancer, which is likely of mechanistic significance.

There is also growing evidence that, when associated with 'paraneoplastic' autoimmune syndromes, cancers are frequently smaller at diagnosis and may have a better prognosis *(25)*. Perhaps the most striking evidence for a potential mechanistic association between autoimmunity and effective anti-cancer immunity comes from studies showing that development of autoimmunity during immunotherapy for a variety of different cancers is a predictor of a better cancer outcome. This may target both (*i*) tissue-specific antigens (e.g. in metastatic melanoma, where development of vitiligo is a predictor of an effective anti-tumor response) *(26)* or (*ii*) non-tissue-specific autoantigens (e.g. systemic high dose adjuvant interferon α-2b therapy for metastatic melanoma inducing new autoantibodies) *(27)*. These associations suggest that autoimmunity can be a byproduct of a successful anti-tumor immune response.

A model of mechanism in systemic autoimmunity

A mechanistic explanation for the association between systemic autoimmunity and cancer has been elegantly explored using autoantibodies specific for myositis, a disease in which skeletal muscle is known to be the target of the autoimmune response. From work done in the Rosen lab, three intriguing observations have led to a compelling model of disease: first, regenerating muscle fibers in myositis patients express high levels of autoantigens whereas the levels are much lower in fully differentiated, healthy muscle; second, antigen expression correlates with myoblast differentiation status in vitro; and finally, cancers associated with myositis also express high levels of these autoantigens whereas the corresponding healthy tissues do not *(13)*. These data suggest that in myositis, an anti-cancer immune response may be misdirected at muscle progenitor cells, resulting in immune mediated injury of regenerating muscle tissue.

Myositis, scleroderma, and other systemic autoimmune diseases are united by many features, including their associations with autoantibodies and cancer, as well as the presence of a self amplifying cycle of tissue destruction and immune system activation with a failure of homeostatic tissue repair. These diseases are also united by the fact that most autoantigens identified to date are ubiquitously expressed proteins, which is incongruent with the highly specific patterns of tissue damage observed in these diseases. Finally, there is a significant subset of patients with characteristic features of systemic autoimmunity that lack any known autoantibodies. These commonalities support the conclusion that undiscovered autoantibodies likely exist and that in addition to their potential clinical utility, identification of the targets of these autoantibodies may lend

significant insights into the underlying mechanisms driving the initiation and propagation of these devastating diseases.

Thesis overview

Despite the identification of numerous disease specific autoantibody specificities, many of which tightly correlate with clinical phenotype and prognosis, the role of these targeted proteins in the initiation and propagation of systemic autoimmunity remains largely elusive. Although some studies have shed insight on a possible role for target tissue progenitor cells as drivers of the immune response through high levels of autoantigen expression, these studies have been limited to myositis, in which the target tissue is well defined and amenable to sampling. In contrast, scleroderma is a disease in which the precise target of the immune response is unknown and a potential role for tissue progenitors in driving the immune response has not been explored.

While most autoantigen identification studies to date have used tractable, immortalized cancer cell lines, these cell types likely express only a small subset of proteins targeted in systemic autoimmune disease. In contrast to previous studies, we propose that the use of appropriately perturbed antigen sources that mimic changes in the target tissue(s) are key to the discovery of novel autoantigens not detected using conventional antigen sources (Figure 1). Moreover, using cells that more closely mimic the true targets of the immune response will likely provide insight into underlying disease mechanisms.

In order to streamline the identification of novel autoantigens, it was necessary to develop a systematic approach to antigen identification that accommodates the

complexity of protein dynamics in perturbed cells while simultaneously limiting the investment of time and financial resources necessary for the identification of any particular target. In order to develop such a system, we initially focused on studies in autoimmune myositis (with known target tissue and stronger understanding of disease mechanism) to develop methods to identify novel antigens from disease relevant antigen sources.

We next sought to apply those methods to the study of scleroderma, a disease in which the target of the immune response is unknown. Based on the disease model that an antitumor immune response may be misdirected against progenitor cells involved in tissue repair, we sought to prove that differentiating progenitor cells are a target of the immune response in scleroderma. The identification of progenitor specific autoantigens in scleroderma may provide important mechanistic understanding and enable the development of disease specific therapies that are desperately needed in this potentially devastating and poorly understood disease.

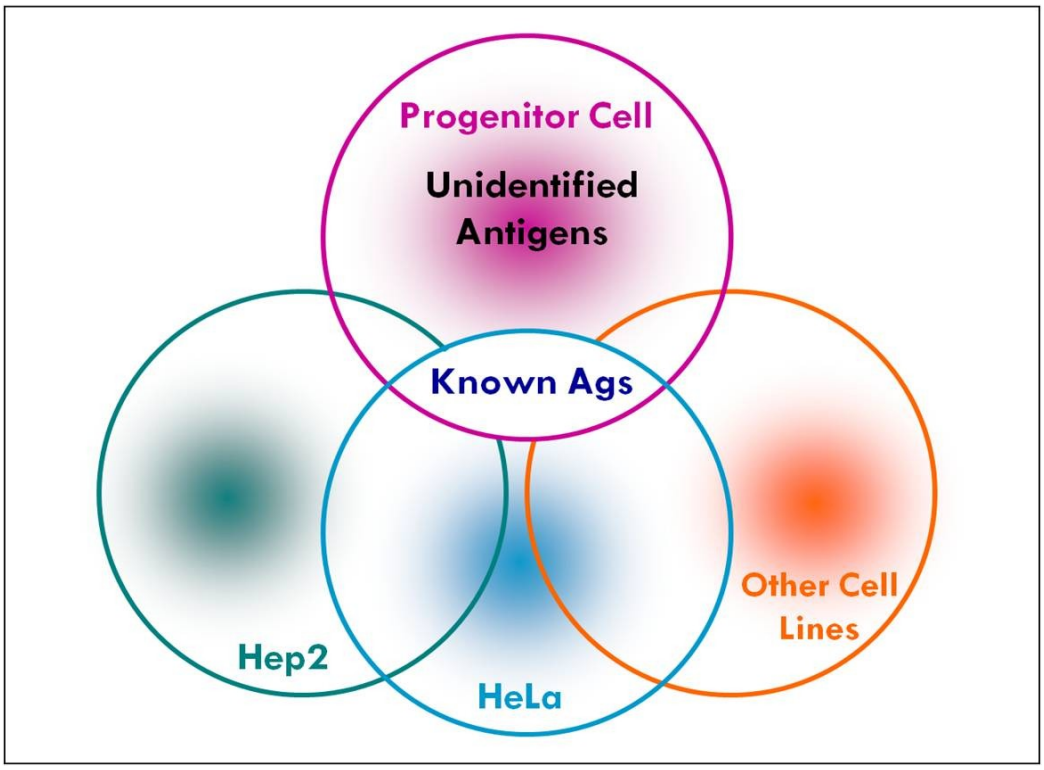


Figure 1: Venn diagram illustrating the importance of protein sources in antigen identification. Previous experiments seeking to identify the targets of autoantibodies have used cell lines such as HeLa (cervical cancer derived) or Hep2 (laryngeal carcinoma derived) cells as antigen source(s), which are tractable, readily available and may differ tremendously from the cell type(s) targeted by the immune response. Currently known autoantigens likely only reflect the small subset of proteins targeted that are expressed in both commonly used cell lines and the true target cell type(s). In contrast, using an appropriately perturbed antigen source (such as a differentiating tissue progenitor) that approximates the cell type(s) targeted by the immune response (pink) will allow the identification of many novel autoantigens and provide significant insight into underlying disease mechanisms.

Chapter 2: Identification of Novel Autoantigens by a Triangulation Approach

Introduction

High titer autoantibodies are frequently observed in systemic autoimmunity, where they are often associated with specific phenotypic manifestations *(28, 29)*. These antibody specificities are therefore useful for diagnosis and prognosis, and provide insights into disease mechanisms. Paradoxically, the vast majority of antigen targets identified to date in systemic autoimmunity are ubiquitously expressed, a finding that is difficult to reconcile with the highly specific patterns of tissue damage observed in these diseases. For example, antibodies against aminoacyl tRNA synthetases are found only in patients with myositis (in which muscle and lung are the primary target organs), and yet these proteins perform an essential function in all cells. Similarly, topoisomerase 1 and centromere proteins are necessarily present in all cell types and are well defined autoantibody targets in systemic sclerosis, a disease in which the skin, lungs, and gastrointestinal tract are most commonly affected. In addition, many patients with systemic autoimmune diseases lack known autoantibody specificities (e.g., in myositis, ~20-40% of patients are autoantibody negative as assessed using currently available clinical assays *(30)*).

In several autoimmune diseases, autoantibodies recognize post-translationally modified proteins, such as citrullinated proteins in rheumatoid arthritis *(10, 11)* and proteins susceptible to granzyme B cleavage *(31)*. It has been proposed that perturbations

that result in the modification or enhanced expression of particular proteins may drive the immune response in some autoimmune diseases.

In the inflammatory myopathies, an increasing body of data supports a role for disease-specific changes in antigen expression within the target tissue in driving the disease-specific immune response. For example, expression of Mi-2 is increased specifically in dermatomyositis (DM) muscle, the disease in which Mi-2 autoantibodies are found *(13)*. Additionally, Mi-2 is expressed at high levels in regenerating muscle cells, but only at low levels in differentiated muscle cells *(32)*. Similarly, 3-hydroxy-3-methylglutaryl-coenzyme A reductase expression is dramatically upregulated by statins both in cultured cells, as well as in muscle from patients with statin-induced necrotizing myopathy (of note, these patients make high titer autoantibodies to this enzyme) *(15)*. Antibodies targeting melanoma differentiation-associated protein 5 (MDA5) have also recently been identified in myositis patients previously considered antibody negative *(33, 34)*. Interestingly, expression of this protein is highly induced by interferon-α (IFN-α), a finding consistent with the strong type I IFN signature that has been well documented in the muscle of patients with DM *(35)*.

The data suggest that screening the appropriately perturbed antigen source, which mimics changes identified in the target tissue in that disease, may enable the discovery of novel autoantibodies. We have therefore focused on autoantibody discovery using antigen sources in which pathways known to be active in the target tissue *in situ* are recapitulated; examples described here are exposure to type I IFN and muscle regeneration. Whereas several methods of antigen identification exist, including screening of recombinant cDNA expression libraries (SEREX), peptide libraries, and

multiple affinity protein profiling, few accommodate the complexity of protein dynamics observed *in vivo*, such as splice variants, post-translational modifications, cleavage, nucleic acid binding, inducible expression, and cryptic epitope exposure.

Serological proteome analysis (SERPA), which combines two-dimensional (2D) electrophoresis with immunoblotting and tandem mass spectrometry *(36)*, is currently one of the most useful techniques to define modified autoantigens. A major advantage of this method is the high sensitivity of both immunoblotting and tandem mass spectrometry, allowing the detection of relatively low abundance antigens within a complex protein mixture *(37)*. This approach is, however, significantly limited by the fact that selection of a low abundance protein spot on a 2D gel by alignment with a patient serum immunoblot is often insufficiently precise. Since numerous candidate proteins are generated by tandem mass spectrometry (MS/MS) sequencing *(38)*, it is critical that candidate antigens are rigorously confirmed to eliminate false positives. This confirmation often requires a substantial investment of time and resources before the correct protein is confirmed.

This chapter describes a rapid multidimensional method of autoantigen discovery using patient sera as probes. Using a combination of (*i*) cells subjected to a perturbation known to be present in the diseased tissue; (*ii*) a modified SERPA strategy that confirms that the relevant spot has been plucked from the gel before sequencing; and (*iii*) mRNA expression profile data that define molecules whose expression is enhanced by the perturbation, enables rapid autoantigen identification. Data is shown that details the identification process for two novel autoantigens whose expression is regulated in the context of two different pathways of pathogenic relevance.

Materials and methods

Human sera

Frozen aliquots of de-identified sera from 141 myositis patients, obtained during routine blood draws at the Johns Hopkins Myositis Center, were used in the preliminary screens for this study. Patients were clinically defined according to the Bohan and Peter criteria *(39, 40)*. Normal sera from 50 healthy donors were also used. All samples were obtained, and studies were performed, in compliance with Institutional Review Board and Health Insurance Portability and Accountability Act regulations.

Cell culture

Human salivary gland (HSG) cells were a generous gift from Dr. Bruce Baum and were cultured in minimal essential medium (MEM; Invitrogen) with 10% heat-inactivated fetal calf serum (HI FCS), 1% L-glutamine, and 1% penicillin/streptomycin at 37°C with 5% CO_2. Human cervical epithelial cells (HeLa) were cultured in Dulbecco's modified eagle medium (DMEM; Invitrogen) low glucose with 10% HI FCS, 1% L-glutamine, and 1% penicillin/streptomycin at 37°C with 7.5% CO_2. Second passage primary human keratinocytes (Lonza) were grown in Keratinocyte Basal Medium-2 (KBM-2) supplemented with the KGM-2 bullet kit (Lonza) without subculturing. For IFN treatment, cells were grown to 70% confluence before being cultured for an additional 24 hours in the absence or presence of 1000 units/mL IFN-α (Sigma). Normal human skeletal muscle cells were cultured and induced to differentiate into myotubes over a two week period as described *(13)*.

Lysate preparation from cultured cells and human muscle biopsies

Culture dishes were kept on ice during two washes with cold phosphate buffered saline (PBS) followed by scraping into lysis buffer (1% Nonidet P-40 (NP40); 20 mM Tris, pH 7.4; 150 mM NaCl; one mM ethylenediaminetetraacetic acid (EDTA)) containing a protease inhibitor cocktail. Human muscle biopsies were obtained from patients seen at the Johns Hopkins Myositis Center. Surgical procedures were performed for patient management, and the research tissue samples were in excess of that required for routine diagnostic purposes. Collection and use of the human biopsy specimens were approved by the Johns Hopkins Institutional Review Board. All patient samples were deidentified, with clinical and laboratory features linked only to the patient code. Detergent extracts were prepared in lysis buffer from frozen muscle biopsies obtained from DM and polymyositis (PM) patients, and from a control individual, as described *(13)*. Patients were defined as having DM or PM if they had probable or definite disease according to the Bohan and Peter criteria *(39, 40)*. All lysates were diluted to working concentrations in gel application buffer (2% SDS; 4% glycerol; 40 mM Tris-HCl, pH 6.8; 0.002% bromophenol blue; 1% β-mercaptoethanol) and boiled for three minutes.

Immunoblotting

Primary immunoblotting screens using lysate pairs from HSG and HeLa cells incubated in the absence or presence of IFN-α or muscle cells at various stages of differentiation, were performed as follows. Five μg/lane of each lysate was electrophoresed on seven cm, 10% SDS-PAGE gels for 170 Vh, then transferred to

nitrocellulose membranes at 100 V for one hour. Membranes were stained with Ponceau (0.3% w/v in 3% trichloroacetic acid) and washed with distilled H_2O to identify molecular weight markers, followed by blocking for one hour at room temperature (RT) in 3% bovine serum albumin in wash buffer (WB) (ten mM Tris-HCl, pH 7.6; 0.5% NP40; 150 mM NaCl). Membranes were then incubated with patient serum diluted in WB (1:5000) for one hour at RT. After extensive washing with WB (five washes for five minutes each), membranes were subsequently incubated with a horseradish peroxidase-conjugated goat anti-human IgG secondary antibody (Jackson Immunoresearch) at 1:10,000 dilution in WB for one hour. Membranes were then washed as above and incubated with enhanced chemiluminescent reagent (Pierce) followed by exposure to X-ray film. For serum M4, a second screen including lysates from HSGs, HeLas, and primary human keratinocytes treated with 1000 units/mL IFN-α was performed as described above.

2-Dimensional (2D) electrophoresis

IFN-α treated cells and muscle cells at day four of differentiation were washed twice with cold PBS followed by two washes with cold 200 mM Sucrose, ten mM Tris HCl, pH 7.4 and lysed directly into isoelectric focusing (IEF) buffer (eight M urea, two M thiourea, 4% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 1% dithiothreitol and protease inhibitors). First dimension focusing was performed using Bio-Rad ReadyStrip immobilized pH gradient (IPG) Strips and Protean IEF Cell per manufacturer's instructions. Seven cm IEF strips were used to determine the optimum pH range for the protein of interest whereas 17 cm IEF strips were used for

sequencing. Briefly, each seven cm IEF strip was loaded with 130 µl consisting of 100 µg lysate, 1.5% 2-hydroxyethyl disulfide, 0.1% bromophenol blue and 0.4% ampholytes (pH dependent on IEF strip pH range). For 17 cm strips, 300 µg lysate was used in a total volume of 300 µl. Up to 12 IEF strips can be run simultaneously; these can subsequently be processed immediately for the second dimension gel (see below), or they can be stored at -80°C. Prior to running the second dimension gels, IEF strips were incubated for 20 minutes at RT in ten mg/ml dithiothreitol, followed by 20 minutes in 25 mg/ml iodoacetamide, both in equilibration buffer (six M urea; 0.375 M Tris-HCl, pH 8.8; 2% SDS; 20% glycerol). IEF strips were then loaded onto 10% SDS PAGE gels (16 cm or seven cm gels, depending on the size of the IPG strips), covered with 1% low melt agarose (Promega) and electrophoresed as described above (16 cm gels were electrophoresed for 850 volt hours, and transferred onto nitrocellulose membranes for four hours at 400 mA). To ensure consistency, IEF strips run in parallel were used for 2D immunoblotting and corresponding protein sequencing experiments.

2D immunoblotting and antigen localization

2D gels were transferred to nitrocellulose membranes, stained with Ponceau to visualize resolved protein spots, and scanned at high resolution for alignment with the subsequent immunoblots. Immunoblots were performed as described. During X-ray film exposure, precise documentation of the membrane perimeter on the film is critical for subsequent alignment of the exposed film with a printout of the Ponceau scan. This enables the immunoblotted spot of interest to be aligned with the corresponding Ponceau-stained protein spot. Once the protein of interest is located on the Ponceau scan, nearby

spot patterns can be used for orientation to find the same spot on the preparative gel for plucking and protein identification.

Protein identification

Preparative SDS-PAGE second dimension gels were made using solutions and procedures that minimize exposure to airborne contaminants. All glassware and equipment were only used for proteomics experiments and were washed twice with 0.1% SDS (w/v) followed by extensive rinsing with ultrapure distilled water and 100% methanol to minimize protein contamination. After performing first and second dimension gel runs as described, gels were fixed in destain (14% v/v acetic acid; 24% v/v methanol) for 15 minutes, incubated for one hour in Gel Code Blue Stain (Thermo Scientific) and then destained in ultrapure distilled water for two to four hours. Destained gels were placed on a light box to facilitate visualization of stained protein spots, and gel plugs containing the protein spot of interest and a nearby "blank" gel plug containing no visible protein were carefully removed from the preparative gel using a sterile pipette tip. The gel plugs were immediately washed twice in 50% methanol for ten minutes each with shaking and stored at -20°C. The remaining preparative gel was washed three times for five minutes each in running buffer (25 mM Tris-HCl; 200 mM glycine; 0.1% (w/v) SDS), and proteins were transferred to nitrocellulose membrane for immunoblotting to verify successful plucking. Gel plugs confirmed to contain the protein of interest were sent to the Johns Hopkins Mass Spectrometry and Proteomics Facility for liquid chromatography tandem mass spectrometry (LC/MS/MS) peptide sequencing.

Microarray

Total RNA was extracted using the Trizol reagent method (Invitrogen) from HSGs cultured for 24 hours in the absence or presence of IFN-α at [1000 units/ml], and from muscle cells at various stages of differentiation (days zero to five). All samples were collected in triplicate. Additional purification was performed on RNeasy columns (Qiagen), and the quality of total RNA samples was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies). Biotin-labeled, complementary RNA (cRNA) was prepared from total RNA according to the chip manufacturer's protocol (Illumina). cRNA was hybridized to Illumina HumanHT12 v3 Expression BeadChips, and signal was detected with streptavidin-Cy3 (IFN treatment). cRNA was hybridized to Illumina Sentrix HumanRef-8 Expression BeadChips, and signal was detected with streptavidin-Cy3 (muscle cell differentiation). All signal intensity quantification was performed using an Illumina BeadStation 500GX Genetic Analysis Systems scanner.

Analytical methods and statistical analysis for microarray data

IFN treatment - A single intensity (expression) value for each Illumina probe on the array was obtained using Illumina BeadStudio software with standard settings and no background correction. The expression values for all the probes for each sample were scaled to have median 256 (2^8) and then log (base 2) transformed. Genes (i.e., Illumina probes) considered to be differentially expressed between two groups of samples were those satisfying the following criteria: (*i*) Welch t-test p-values less than or equal to 0.01 (10^{-2}) *(41)*; (*ii*) a Benjamini-Hochberg false discovery rate (FDR) less than or equal to 0.25 *(42)*; (*iii*) a fold change above 2.0 (calculated using geometric means); and (*iv*) the

expression value of the gene is above the Illumina BeadStudio calculated background for all three samples in the group.

Muscle cell differentiation - Preliminary analysis of the scanned data was performed using Illumina BeadStudio software. Any gene consistently below a background threshold level of D=.98 for all samples was eliminated from further analysis. Bead averaged data were normalized by Z transformation. Briefly, microarray data were $\log_{10}$ transformed and data were standardized to the mean expression of the dataset and represented in terms of a Z score.

Antigen identity confirmation

Embryonic myosin light chain 4 (MYL4) cDNA was cloned from myoblasts at day four of differentiation, and was sequence verified. Interferon-induced protein with tetratricopeptide repeats 3 (IFIT3) cDNA (Origene) and MYL4 cDNA were transiently transfected into HeLa cells using Lipofectamine (Invitrogen), per the manufacturer's protocol.

Mouse monoclonal anti-IFIT3 and anti-MYL4 antibodies were obtained from BD Transduction Labs and Novus, respectively. Cell lysates were transferred onto nitrocellulose membranes and blocked in Tris buffered saline/Tween 20 (TBST; ten mM Tris-HCl, pH 7.4; 150 mM NaCl; 0.05% (v/v) Tween 20) with 5% (w/v) powdered milk for one hour at RT. Membranes were then incubated with anti-IFIT3 or anti-MYL4 antibodies (diluted 1:3,000 or 1:1,000, respectively, in TBST/5% milk) overnight at 4°C. Membranes were washed five times for five minutes each with TBST/5% milk at RT, followed by incubation with a 1:10,000 dilution of horseradish peroxidase conjugated

anti-mouse secondary antibody in TBST/5% milk for one hour at RT. After washing five times for five minutes each using TBST buffer, membranes were incubated for five minutes with enhanced chemiluminescent reagent (Pierce) and then exposed to X-ray film.

^{35}S-methionine-labeled IFIT3 protein was synthesized using the TNT T7 Coupled Reticulocyte Lysate System (Promega) for *in vitro* transcription and translation (IVTT) using IFIT3 cDNA per the manufacturer's instructions. Immunoprecipitations using ^{35}S-methionine-labeled IFIT3 protein were performed as described, with all steps carried out at 4^{o}C *(33)*. Briefly, one µl of IVTT product was diluted in lysis buffer with protease inhibitors, one µl of patient serum was added and the mixture was rotated for one hour. 35 µl of immobilized Protein A agarose (Pierce) was then added, and the mixture was rocked for 20 minutes. Immunoprecipitates were subsequently washed four times with lysis buffer, electrophoresed on 10% SDS-PAGE gels and visualized by fluorography.

Results

Detection of IFN-α induced autoantigens, and selection of a novel specificity

To screen for new autoantibodies targeting IFN-α-induced proteins, lysates made from untreated and IFN-α treated cells were immunoblotted with myositis patient sera. A representative screening blot performed on untreated and IFN- α treated HSGs is shown (Fig. 2A). Several different patterns were observed: some sera immunoblotted proteins present exclusively in IFN-α treated cell lysates (e.g., the 60 kDa protein blotted by M4); others blotted proteins present at baseline, but upregulated by IFN-α (e.g., the 52 kDa protein blotted by M1 and M3); some blotted proteins whose expression remained unchanged (e.g., M2 and M5) and some did not blot anything (e.g., M6). We focused on sera that detected novel specificities induced by IFN-α treatment; hence, of the DM patient sera shown in Fig 2A, only M4 was of interest (the 52 kDa band blotted by M1 and M3 is Ro52 (TRIM21), an autoantigen which is known to be targeted in patients with myositis). Of note, we found antibodies recognizing IFN-α induced proteins in 16/141 (11%) of the myositis sera; IFN-inducible specificities were never detected with healthy control sera (data not shown).

A subsequent screen was performed to determine if the 60 kDa antigen blotted by serum M4 could be detected in other cell types treated with IFN-α. As shown in Fig 2B, the 60 kDa protein was also detected in HeLa cells and primary human keratinocytes upon treatment with IFN-α, suggesting that local type I IFN signaling may induce autoantigen expression by a variety of cell types. Moreover, this enabled the use of data from multiple IFN-treated cell types for antigen identification.

Antigen identification strategy

We next pursued the identification of the 60 kDa protein using 2D gel electrophoresis and LC/MS/MS peptide sequencing. A critical aspect of this methodology is precise localization of the protein of interest by immunoblotting a replicate 2D gel (Fig. 3A). It is important to scan the Ponceau stained membrane prior to immunoblotting because careful alignment of the immunoblot with a printout of the scanned Ponceau image allows identification of the antigenic protein spot to be plucked. The surrounding pattern of spots visible only on the Ponceau scan is critical for locating the spot of interest on an identically run preparative 2D gel stained with colloidal blue (Fig. 3B). Of note, spot identification can be a challenging step if the antigen of interest is relatively low abundance (hence faintly stained, as highlighted for the M4 antigen in Fig. 3B), which may result in inaccurate plucking. Once removed, the gel plug is sent for peptide sequencing by LC/MS/MS and a list of candidate proteins is generated by BLAST searching all of the peptide sequences generated against a mammalian protein sequence database (MASCOT). Although only a tiny gel plug containing the spot of interest is used for sequencing, the high sensitivity of the LC/MS/MS frequently results in lists of ~10-20 candidate proteins from the detected peptides, even after subtracting contaminants sequenced from a nearby "blank" gel plug. Pursuing candidate lists obtained from inaccurate spot selections therefore wastes significant time and resources to eliminate spurious candidates.

We therefore developed a rapid, inexpensive quality control step to confirm that the correct protein spot is plucked before sequencing, and have confirmed that this

dramatically improves the efficiency and accuracy of antigen identification. This is accomplished by transferring the proteins remaining after plucking to a nitrocellulose membrane, and immunoblotting with the patient serum. If the spot of interest is accurately plucked, there will be a halo of signal around the removed spot (as is shown in the boxed area, Fig. 3C). In those cases where it is challenging to accurately identify which spot to pluck, we recommend performing several plucks (including multiple gels and a variety of cell types when possible), and performing a post-pluck immunoblot on each. This strategy rapidly and inexpensively identifies the best gel plug, and only this is then sent for sequencing. LC/MS/MS peptide sequencing of a confirmed pluck of the M4 antigen from IFN-α treated HeLa lysate resulted in a list of twelve candidate proteins in the 50-75 kDa range that were absent from a nearby "blank" gel plug (Table 1). As expected, most of these proteins also had a predicted isoelectric point (pI) similar to that observed for the antigen of interest (estimated to be ~5.2 from the 2D immunoblot).

The triangulation approach requires at least one additional piece of information about the protein of interest to allow multiple different paths of pursuit to converge on the correct antigen identity. Fortuitously, in the case of antigen M4, we were able to take advantage of expression profiling data from IFN-α treated cells performed for an unrelated project. We compared the list generated by proteomic sequencing of the antigen from IFN-α treated HeLa lysate (Table 1) to a list of genes that were induced by at least two-fold in IFN-α treated HSGs (Table 2). This analysis quickly enabled us to focus on a single leading candidate present on both lists: IFN-induced protein with tetratricopeptide repeats 3 (IFIT3).

Confirmation of antigen identity

Several strategies were used to definitively demonstrate that the protein targeted by high titer serum autoantibodies of patient M4 is IFIT3. (*i*) Immunoblots performed with a monoclonal anti-IFIT3 antibody confirmed that IFIT3 is strongly induced in IFN-α treated HeLa cells and that it comigrates with the 60 kDa M4 protein (Fig. 4A, lanes 1, 2, 5 & 6); (*ii*) Both the M4 patient serum and the anti-IFIT3 monoclonal antibody immunoblot a 60 kDa protein expressed in HeLa cells transiently transfected with IFIT3 cDNA (Fig. 4A, lanes 4 & 8), but absent in untransfected cells (lanes 3 & 7); and (*iii*) ^{35}S-methionine labeled IFIT3 (^{35}S-IFIT3) was generated by IVTT using IFIT3 cDNA. Serum M4 immunoprecipitated ^{35}S-IFIT3 (Fig. 4B), whereas other myositis patient sera known to be anti-IFIT3 antibody negative (as assessed by immunoblot of IFN-α induced cell lysates, data not shown), did not. Healthy control sera also did not immunoprecipitate ^{35}S-methionine-IFIT3 (data not shown).

Validation of the above approach using muscle cells at successive stages of differentiation to screen for new myositis autoantibodies: identification of MYL4

We sought to validate the approach described above by using similar methods to detect myositis autoantigens whose expression is enhanced in the context of additional pathways active in diseased tissue. Specifically, since myositis autoantigen expression has been demonstrated to be enhanced in differentiating muscle cells *(13)*, we used these as a novel source for autoantigen discovery. Myositis serum S293 immunoblotted a 28 kDa band that was robustly expressed in muscle cells at day four (or thereafter) of differentiation, but was absent from myoblasts (day zero cultures), as well as HeLa, HSG

and keratinocyte lysates (Fig. 5A and data not shown). Lysates made from myoblasts harvested at day four of differentiation were electrophoresed on 2D gels (1st dimension, pI 3-10; 2nd dimension, 12% SDS-PAGE). Duplicate runs were done: the first gel was transferred to nitrocellulose and used for immunoblotting with serum S-293 to identify the spot corresponding to the 28 kDa protein. The second, replicate gel was stained with colloidal coomassie blue immediately after SDS-PAGE, and the 28 kDa spot was plucked and processed for peptide sequencing, after confirming by post-pluck blotting that the correct spot had been plucked

To aid in identifying the real target autoantigen from the resulting list of candidate proteins, we used microarrays to obtain gene expression patterns during myoblast differentiation (days zero to five). In the case of the unidentified 28 kDa antigen, there were eight candidate proteins from the 2D proteomic analysis. Two of the top candidates were known muscle-specific proteins; embryonic myosin light chain 4 (MYL4) and troponin I, skeletal slow twitch (TNNI). When these were searched in our myoblast differentiation gene expression database, only one was absent at day zero, but was dramatically increased after three to four days of differentiation. This antigen was MYL4, a muscle-specific, differentiation stage-specific 28 kDa protein (Fig. 5B), making it the top candidate.

We validated that serum S293 does have antibodies to MYL4 in the following ways. Immunoblots performed using a monoclonal anti-MYL4 antibody confirmed that the kinetics of expression during myoblast differentiation were identical to those seen using S293 serum (Fig. 5C). Furthermore, the 28 kDa band blotted by both the monoclonal antibody and S293 co-migrated (Fig. 5D). Lastly, we cloned the cDNA for

MYL4 from day four myoblast cultures and transiently transfected it into HeLa cells. The 28 kDa band was detected by immunoblotting in transfected lysates only, using both the monoclonal antibody and serum S-293 (Fig 5D). Taken together, the above approaches confirm that the 28 kDa antigen blotted by S293 is MYL4.

Expression of IFIT3 and MYL4 in muscle biopsy lysates

We used lysates made from muscle biopsies (the target tissue in myositis) to examine expression of IFIT3 and MYL4 *in vivo*. Lysates were made from control, DM and polymyositis (PM) muscle. IFIT3 was detected in DM muscle only, and was absent from PM and control muscle (Fig. 6, lanes 4-6), consistent with the well-documented type 1 IFN signature in DM muscle *(43)*. Of note, IFIT3 expression was also found to be enhanced (though at lower levels) during muscle differentiation (absent at day zero, present at day four, Fig 6 lanes 1 & 3). MYL4 was robustly expressed in both DM and PM muscle biopsy lysates but was not detected in control muscle (Fig. 6). This expression pattern is consistent with immunohistochemical data showing extensive muscle cell regeneration in both DM and PM *(13)*, but not in control muscle.

Discussion

The triangulation approach described here has three main features that improve the efficiency of autoantigen identification (Fig. 7): (*i*) Use of antigen sources that express pathways shown to be active in the diseased tissue *in vivo* (e.g., specific cell or tissue types, differentiation stages, activated signaling pathways, etc). This may allow the detection of novel autoantigens of mechanistic importance missed by other screening strategies; (*ii*) Confirmation of spot selection by post-pluck immunoblotting prior to LC/MS/MS peptide sequencing; and (*iii*) Use of data from additional methods (e.g. biochemical properties of the antigen, microarray analysis of gene expression) to filter the proteomic data, prior to rigorous experimental confirmation of antigen identity. This filtering step narrows the number of candidates and limits the expense and time required for confirmation. As used in this study, an antibody specific for the protein candidate and/or a cDNA expression plasmid are very useful in providing unequivocal evidence of antigen identity. Additional methods of antigen identity confirmation may include knocking down expression of the protein with specific siRNAs, screening a knockout cell line, comparison of protease specific cleavage patterns, or 2D immunoblotting to confirm exact co-localization using the patient serum and a protein specific antibody.

A potential disadvantage of this strategy is that identification of proteins by 2D gel and LC/MS/MS sequencing can be challenging for very large and poorly soluble proteins *(37)*, such as those in cell membranes, although sequencing of cleavage fragments instead of the full length protein may circumvent this limitation. Post-translational modifications also may pose significant challenge, as the modified residues can result in peptide fragments that are difficult to identify, especially in the case of

glycosylation *(38)*. Furthermore, filtering by changes in gene expression is not suited to autoantigens generated by post-translational modification. Evaluating novel antigens for post-translational modifications prior to sequencing permits the identification of modifications that should be included in the analysis of the LC/MS/MS data and potentially allows for their removal prior to sequencing. Finally, if a novel antigen is a protein that has not been previously identified, it may be absent from the mammalian database and thus may also be missed by this approach.

These studies emphasize that current technologies enable rapid and efficient autoantigen discovery, and allow novel autoantigens expressed in pathways relevant to disease pathogenesis to be defined. Such autoantigen targets – when induced by immune effector or tissue repair pathways - may be of particular importance in amplifying and sustaining injury in these diseases. These molecules may provide a new class of probes for diagnosis and monitoring.

Table 1: Candidate proteins generated by LC/MS/MS peptide sequencing of a 2D gel spot containing the 60 kDa M4 antigen, listed in order of strength and restricted to those with a predicted MW of 50-75 kDa.

GI Number	Protein[a]	MW (kDa)[b]	pI[b]	Gene Symbol
gi|31542947	60 kDa heat shock protein, mitochondrial	61.1	5.7	HSPD1
gi|1777782	Interferon-induced protein with tetratricopeptide repeats 3	55.9	5.1	IFIT3
gi|28317	Keratin 10	58.8	5.1	KRT10
gi|20070228	Nucleobindin 1	53.9	5.1	NUCB1
gi|35655	Protein Disulfide Isomerase	57.1	4.8	P4HB
gi|624873	Rab GDI alpha	50.4	5.2	GDI1
gi|547754	Cytokeratin-2e	65.9	8.1	KRT2
gi|2343085	Programmed cell death protein 4	51.7	5.0	PDCD4
gi|12803709	Keratin 14	51.7	5.1	KRT14
gi|1136741	T complex protein 1 subunit theta	58.6	5.8	CCT8
gi|2160719	Myc box-dependent-interacting protein 1	64.7	5.0	BIN1
gi|121570	78 kDa glucose-regulated protein	72.3	5.1	HSPA5

[a] Only proteins predicted to be 50-75 kDa are shown.

[b] Predicted based on amino acid sequence.

GI = GenInfo Identifier; aa = amino acids; MW = molecular weight; kDa = kilodaltons; pI = isoelectric point

Table 2: Genes whose transcripts are induced by ≥2 fold with IFN-α treatment of human salivary gland (HSG) cells with predicted MW of 50-75 kDa.

Gene Symbol	Fold Induction	Length (aa)	MW (kDa)[a]
IFI44	28.2	444	50.5
IFI44L	30.1	452	51.3
BCL2L13	2.04	485	52.7
WARS	3.45	471	53.2
TRIM38	2.52	465	53.4
TRIM21	5.97	475	54.2
IRF7	13.73	503	54.3
MUC13	2.56	511	54.5
IFIT2	12.88	472	54.6
SERPING1	2.2	500	55.2
IFIT1	27	478	55.4
IFIT5	2.43	482	55.8
IFIT3	31.2	490	56.0
LAP3	3.94	519	56.2
TRIM5	3.46	493	56.3
TRIM22	8.9	498	56.9
TRIM69	2.9	500	57.4
BTN3A1	2.09	513	57.7
OASL	8.33	514	59.2
EIF2AK2	3.34	551	62.1
REC8	2.25	547	62.6
SLC15A3	6.93	581	63.6
BTN3A3	2.09	584	65.0
LGALS3BP	4.04	585	65.3
UNC93B1	2.33	597	66.6
GBP1	4.16	592	67.9
TRIM25	2.13	630	71.0

[a] Predicted based on amino acid sequence.

aa = amino acids; MW = molecular weight; kDa = kilodaltons

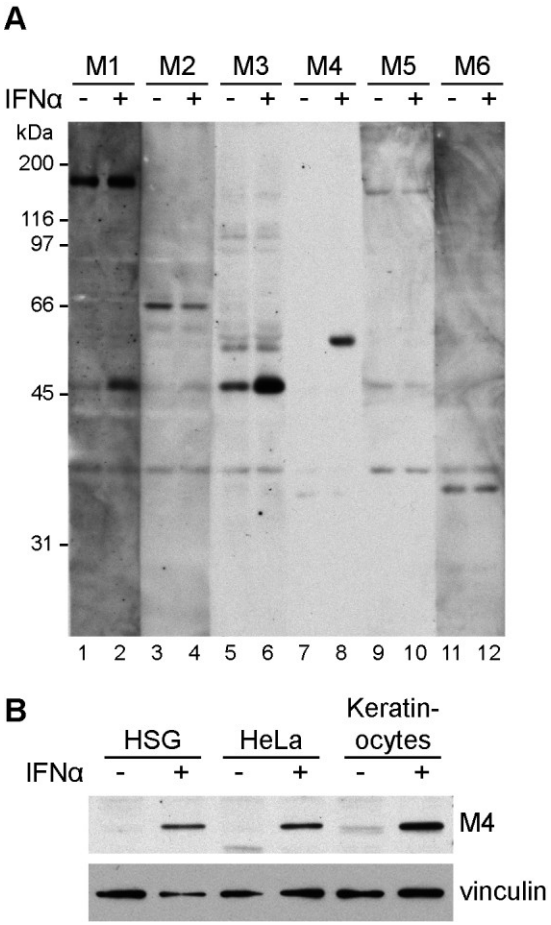


Figure 2: Patient sera detect novel myositis autoantigens in IFN-α treated cells. (A) Immunoblot screen of untreated and IFN-α treated HSG cell lysates with six myositis patient sera (M1-M6). Note that patient M4 has high titer autoantibodies targeting a 60 kDa protein only detectable after IFN-α treatment (lanes 7 & 8). Migration of molecular weight standards are shown down the left side. (B) IFN-α treatment induces expression of the 60 kDa M4 antigen in HSG, HeLa and primary keratinocyte cultures. Vinculin blot reflects the amount of protein loaded in each lane.

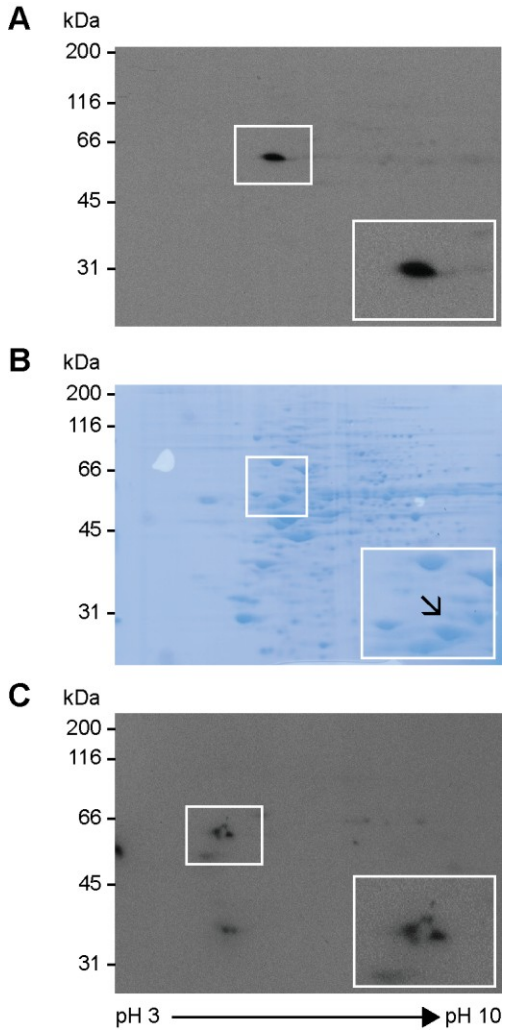


Figure 3: Immunoblot confirmation of protein spot selection increases antigen identification efficiency. (A) 2D immunoblot of IFN-α treated lysate using patient serum M4. First dimension: pI 3-10 IEF; second dimension, 10% SDS-PAGE. The area outlined by the white box is magnified in the lower right corner of each image. (B) Identical replicate preparative gel stained with gel code blue. The spot to be plucked (marked with the arrow) is identified by aligning the Ponceau stained transfer and the immunoblot obtained in A. (C) Immunoblot confirming accurate removal of a gel plug containing the antigen of interest from the preparative gel. Note the halo of signal visible around the removed gel plug.

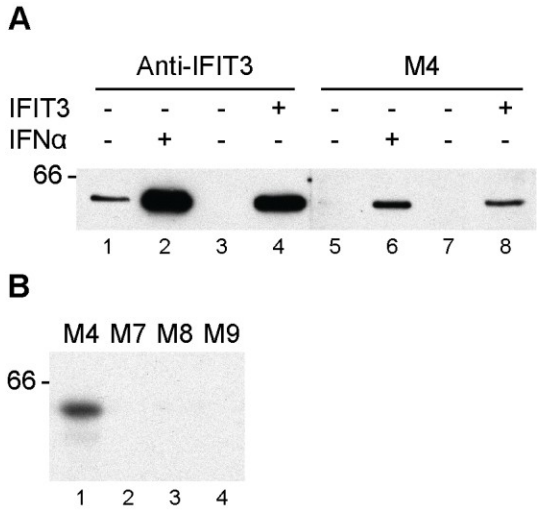


Figure 4: Confirmation that the 60 kDa M4 antigen is IFIT3. (A) Immunoblots were performed with lysates made from HeLa cells incubated in the absence (lanes 1 & 5) or presence (lanes 2 & 6) of IFN-α, or transiently transfected with IFIT3 cDNA (lanes 4 & 8) or vector only (lanes 3 & 7). Equal amounts of protein were loaded in each gel lane. Immunoblots were performed with monoclonal anti-IFIT3 antibody (lanes 1-4) or M4 serum (lanes 5-8). (B) ^{35}S-methionine labeled IFIT3 is immunoprecipitated by the M4 serum (lane 1), but not anti-IFIT3 antibody negative myositis sera (M7-M9, lanes 2-4).

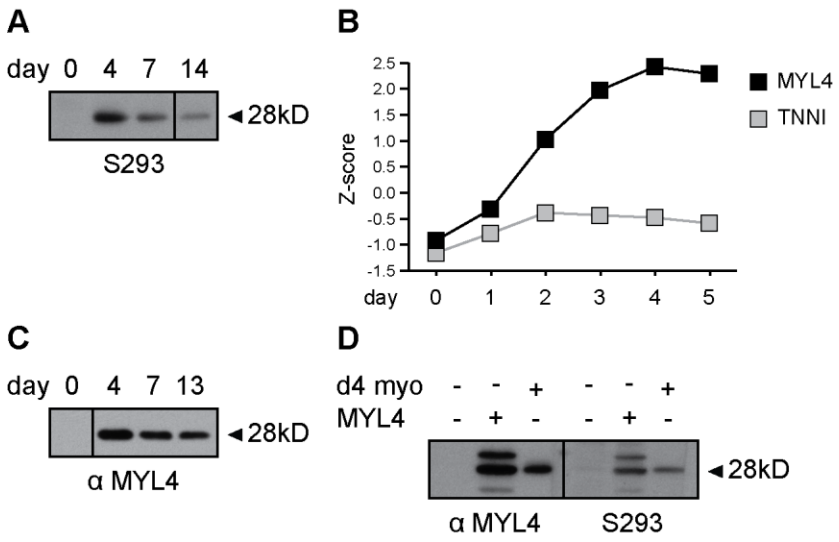


Figure 5: Confirmation that the 28 kDa antigen blotted by S293 serum is MYL4. (A & C) Immunoblots were performed with lysates made from muscle cells harvested at various times during differentiation, using serum S293 (panel A) and monoclonal anti-MYL4 antibody (panel C). Equal amounts of lysate were electrophoresed in each gel lane. (B) Gene expression profiles of muscle cells harvested at days zero through five of differentiation show that MYL4 gene expression is absent at day zero, but is robust by day four, a pattern consistent with the protein expression in the same lysates. TNNI gene expression remained low throughout differentiation. Gene expression values are represented as a Z-score. (D) Untransfected HeLa cell lysates (denoted "MYL4 -"), or HeLa cells transiently transfected with MYL4 cDNA (denoted "MYL4 +", middle lane of each set), or muscle cells at day four of differentiation were immunoblotted with monoclonal anti-MYL4 antibody or S293 serum. MYL4 (28 kDa) was detected in transfected HeLa cells and the day four muscle cell lysates by both. Equal protein amounts were electrophoresed in all gel lanes. (A, C & D) Vertical lines between lanes indicate where lanes on the original autorads have been spliced out for clarity of presentation.

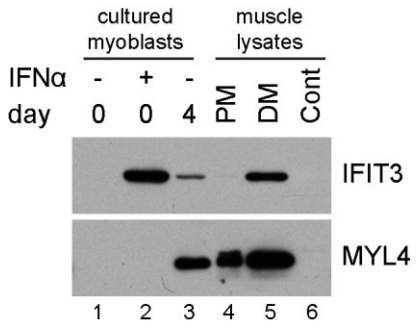


Figure 6: IFIT3 and MYL4 are detected in myositis muscle biopsy lysates, but not in control muscle. Lysates made from undifferentiated myoblasts (day zero) incubated in the absence or presence of IFN-α (lanes 1 & 2), muscle cells at day four of differentiation (lane 3) and lysed muscle biopsies from patients with PM, DM or control individuals (lanes 3-6) were immunoblotted with monoclonal antibodies against IFIT3 or MYL4. Equal amounts of protein were loaded in each gel lane.

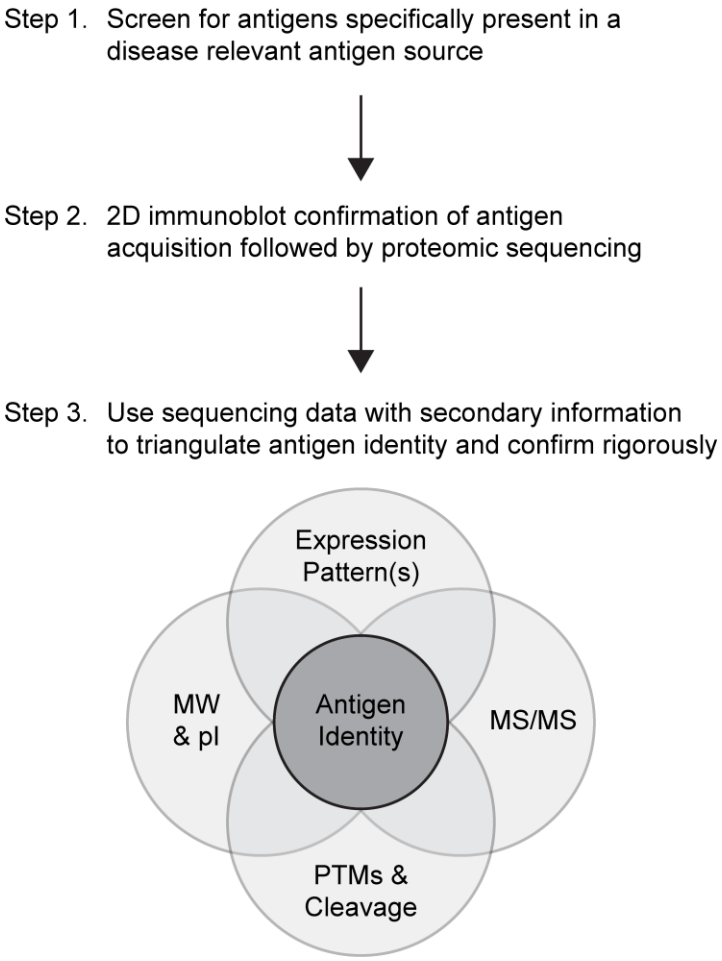


Figure 7: Summary of the triangulation approach to antigen identification. This strategy can be adapted for a variety of antigen identification investigations. Efficient antigen identification is facilitated by focusing on a disease relevant antigen source, confirming accurate plucking of the protein spot of interest by immunoblotting the remaining 2D gel, and combining multiple pieces of information about the antigen to efficiently narrow the candidate pool prior to rigorous confirmation of antigen identity.

Chapter 3: Stem Cells are Highly Enriched Sources of Scleroderma Autoantigens

Introduction

Scleroderma, also known as systemic sclerosis, is a heterogeneous systemic autoimmune disease characterized by immune activation, vasculopathy, and fibrosis. Biopsies of early diseased tissue show damage to small vessels, particularly arterioles, with signs of endothelial cell activation and apoptosis, smooth muscle cell and pericyte proliferation, and fibroblast activation *(44)*. Although this vasculopathy results in tissue hypoxia and the upregulation of numerous angiogenic factors, including vascular endothelial growth factor (VEGF), scleroderma patients exhibit severe impairments in their ability to repair and replace damaged vessels *(45-47)*. Ultimately, this pathology manifests as an array of clinical phenotypes that can include skin fibrosis, esophageal dysfunction, interstitial lung fibrosis, pulmonary arterial hypertension, and renal crisis *(48)*.

Although the mechanism of disease initiation and propagation remains a mystery, several lines of evidence suggest a prominent role for the immune system in scleroderma pathogenesis. Patient biopsies show that early pathogenic changes in small blood vessels are accompanied by pronounced perivascular mononuclear immune cell infiltration consisting primarily of T and B cells, macrophages, and mast cells *(49)*. There are also reports of associations between particular HLA class II molecules and scleroderma clinical phenotypes *(50, 51)*, increased serum levels of Th2 cell-derived cytokines *(52)*,

and oligoclonal expansion of T cells in scleroderma skin lesions *(53)* – all suggesting that the immune response in scleroderma is antigen specific. The strongest evidence implicating the immune system in scleroderma pathogenesis is the production of high titer, class-switched disease specific autoantibodies, which are used for diagnosis and prognosis of the disease *(48)*.

Autoantibodies in scleroderma

Anti-nuclear autoantibodies (ANAs) are found in ~90% of scleroderma patients and specific antibodies are often associated with specific clinical phenotypes *(49)*. For example, anti-centromere antibodies (ACA) are associated with limited cutaneous scleroderma and an increased risk for calcinosis, digital loss, and pulmonary hypertension. In contrast, anti-topoisomerase 1 antibodies (anti-topo 1) are very specific for patients with diffuse cutaneous scleroderma and pulmonary fibrosis. A third example is the association of antibodies against RNA polymerase III with severe skin sclerosis and an increased risk of renal crisis *(54)*.

Although these antibodies are clinically useful, their role in the pathogenesis of scleroderma remains unclear. For instance, antibody titers do not strictly correlate with disease severity and several autoantibodies have been identified in multiple different autoimmune diseases, suggesting that the antibodies themselves are unlikely to be pathogenic. Target antigens identified to date are mostly intracellular, ubiquitously expressed proteins, which is incongruent with the tissue-specific patterns of damage observed *(28, 29, 55)*. Finally, confusion regarding the role of autoantibodies in scleroderma is confounded by the ~10% of patients with scleroderma in the absence of

any known autoantibodies. Whereas the clinical features of these patients suggest an autoimmune mechanism similar to that of the majority of scleroderma patients, speculation remains about whether or not autoantibody negative scleroderma might constitute an entirely different disease process. Although current evidence does not point to a pathogenic role for autoantibodies in scleroderma, their tight associations with clinical phenotypes suggest that these antibodies may mark the specificities of cytotoxic T cells and other cell mediated responses playing an important role in pathogenesis.

Cancer and scleroderma

As in other systemic autoimmune diseases, scleroderma is associated with an increased risk of cancer. For scleroderma patients, lung cancer is the most common malignancy, but there is also a striking temporal association in some cases between the diagnoses of breast cancer and scleroderma *(23, 24)*. Interestingly, in a study of 23 scleroderma patients who had also been diagnosed with cancer, a tight temporal relationship between malignancy diagnosis and the onset of scleroderma was found in patients positive for anti-RNA Pol I/III antibodies, but this relationship was not seen in patients positive for anti-topoisomerase 1 and anti-centromere antibodies. Moreover, the tumors from the scleroderma patients positive for anti-RNA Pol I/III antibodies were found to highly express RNA Pol III whereas tumors from scleroderma patients negative for this antibody and healthy tissue controls did not *(56)*. Recent work has extended this observation by demonstrating the presence of either mutations or loss of heterozygosity of the *POLR3A* gene encoding the RPC1 subunit of RNA polymerase III in the cancers of some of these patients with scleroderma and anti-RNA Pol III antibodies *(57)*.

Moreover, for a subset of these patients, CD4 T cells reactive to peptides containing the patient specific mutations as well as wild type RPC1 peptides were detectable, suggesting that a somatic mutation within a cancer can result in the maturation of autoreactive lymphocytes and high titer autoantibodies specifically associated with scleroderma.

Progenitor cells may link cancer and systemic autoimmune disease

Several observations made by the Rosen lab in the study of another systemic autoimmune disease, autoimmune myositis, have led a compelling model of systemic autoimmunity: first, regenerating muscle fibers in myositis patients express high levels of autoantigens whereas the levels are much lower in fully differentiated, healthy muscle; second, antigen expression correlates with myoblast differentiation status in vitro; and finally, cancers associated with myositis also express high levels of these autoantigens while the corresponding healthy tissues do not *(13)*. These data suggest that in myositis, an anti-cancer immune response may be misdirected at muscle progenitor cells, resulting in immune mediated injury of regenerating muscle tissue. The shared features of systemic autoimmune diseases, including their associations with autoantibodies and cancer as well as the presence of a self-amplifying cycle of tissue destruction and immune system activation with a failure of homeostatic tissue repair, lead us to suspect that this model may be applicable to other systemic autoimmune diseases, including scleroderma.

The hypothesis that an anti-tumor immune response may target proteins also expressed by progenitor cells is supported by the phenotypic overlap between cancer cells and tissue progenitors. Shared functions, including cellular homing and migration, are

critical in the processes of development, tissue repair, and cancer metastasis, all of which utilize similar mechanisms to target stem cells to their respective destinations *(58)*. Additionally, two independent studies founds embryonic stem cell-like (ESC-like) expression signatures in human cancers, one showing that the ESC-like signature in epithelial cancers strongly predicted metastasis and death *(59)* while the other found ESC genes are more frequently overexpressed in several types of poorly differentiated tumors *(60)*. Whereas the connection between stem cells and cancer is an area of ongoing investigation *(61)*, current evidence highlighting the similarities in cell phenotypes and gene expression patterns lend support to the idea that an antitumor immune response may target molecules also expressed by tissue progenitor cells.

A model of mechanism in scleroderma

In scleroderma, we hypothesized that an anti-tumor immune response is misdirected at target tissue progenitor cells, possibly those involved in the repair of damaged blood vessels (Figure 8). Historically, experiments seeking to identify the targets of scleroderma autoantibodies have used tractable cell lines such as HeLa (cervical cancer derived) or Hep2 (laryngeal carcinoma derived) cells as the antigen source (Figure 1). In contrast, we propose that by using antigen sources that approximate the target of the immune response, we will discover novel and mechanistically relevant autoantigens.

Defining a disease relevant antigen source in scleroderma

Stem cells are distinguished from other cell types by unlimited self renewal

capabilities and the capacity to differentiate into more specialized progenitors and cell lineages. Resident stem cells contribute to the homeostasis and regeneration of tissues derived from all three germ layers. Bone marrow derived stem cells are also able to migrate to distant tissues where local niches may specify differentiation into tissue specific progenitors *(62)*. Because the target of the immune response in scleroderma is unknown, this investigation started with undifferentiated human embryonic stem cells (ES), which are capable of differentiating into any cell type in the human body. When these cells are cultured in suspension in the absence of anti-differentiation factors, they spontaneously differentiate and form three-dimensional aggregates called embryoid bodies (EBs) *(63, 64)*. These EBs are heterogeneous, but contain tissues from all three germ layers that are spatially patterned and grossly recapitulate the earliest stages of embryonic development *(65)*. Since the differentiation of these embryoid bodies is not directed down any particular path, they should provide an unbiased source of proteins expressed by many cell types at many stages of differentiation, likely enabling the detection of relevant scleroderma progenitor antigens.

In this chapter, we demonstrate that novel scleroderma autoantigens are detected in lysates of differentiating human embryonic stem cells using scleroderma patient sera as probes. We also demonstrate that these differentiating stem cells are particularly enriched for proteins targeted by autoantibodies in patients previously considered autoantibody negative. Further, we were able to identify keratin 8 and fetuin A as two novel autoantigens targeted by "autoantibody negative" patients. Finally, we demonstrate that immunofluorescence of embryoid bodies can be used to colocalize the expression of autoantigens and lineage markers to aid in the identification of specific cell lineages

expressing high levels of autoantigens, potentially allowing the identification of the true target(s) of the autoimmune response in scleroderma.

Materials and methods

Human sera

Frozen aliquots of de-identified sera from 147 scleroderma patients were obtained during routine blood draws at the Johns Hopkins University Scleroderma Center after written informed consent. All patients either met American College of Rheumatology criteria for SSc *(66)*, or had at least three of five features of CREST syndrome (calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, telangectasias), or had definite Raynaud's phenomenon, abnormal scleroderma pattern nail fold capillaries, and the presence of a scleroderma specific autoantibody (anti-topo 1, ACA or anti-RNA polymerase III). Normal sera from 17 healthy donors were also used. All samples were obtained, and studies were performed, in compliance with Institutional Review Board and Health Insurance Portability and Accountability Act regulations.

Cell culture

Human embryonic stem cell (ES) culture and embryoid body differentiation were performed as previously described *(67)*. Briefly, ES cells (WA-09; passage 39) were cultured on mouse embryonic fibroblasts (MEFs) in Dulbecco's modified Eagle's medium (DMEM)/F12 (GIBCO), supplemented with 20% KnockOut Serum Replacement (GIBCO), 0.1 mM b-mercaptoethanol, and six ng/ml of FGF-2. Cells were passaged with six U/ml of dispase in ES media, washed, and replated at a dilution of 1:3–1:5. For embryoid body differentiation, undifferentiated ES colonies were resuspended in ES medium/20% fetal calf serum without bFGF and plated on poly-HEMA coated

plates to prevent attachment. Cystic EBs emerged after five days in suspension and on day ten, ~80% of EBs were cystic.

HeLa cells were cultured in Dulbecco's modified eagle medium (DMEM; Invitrogen) low glucose with 10% HI FCS, 1% L-glutamine, and 1% penicillin/streptomycin at 37°C with 7.5% CO_2. K562 cells were cultured in RPMI supplemented with two mM glutamine and 10% FBS at 37°C in 5% CO_2. U2OS cells were cultured in McCoy's 5A media supplemented with 1% L-glutamine, 10% HI FCS, and 1% penicillin/streptomycin at 37°C in 5% CO_2. HEK293 cells were cultured in DMEM (Invitrogen) high glucose with 10% HI FCS, 1% L-glutamine, and 1% penicillin/streptomycin at 37°C with 5% CO_2. Jurkat cells were cultured in RPMI 1640 (Gibco/Invitrogen), 10% FCS, ten mM Hepes, 1% penicillin/streptomycin in 5% CO_2 at 37°C.

Lysate preparation

All steps of lysate preparation for all cell types were performed on ice. For adherent cells (including undifferentiated ES, HeLa, U2OS, HEK293): cells were washed with cold PBS (twice, with ten ml per plate) prior to being scraped into lysis buffer (NP40LB) (1% Nonident P40 (NP40); 20 mM Tris pH 7.4; 150 mM NaCl; one mM EDTA) containing a protease inhibitor cocktail (PIs) (pepstatin, antipain, leupeptin, and phenylmethylsulfonyl fluoride) using approximately one ml per every two to three plates, depending on cell density. All lysates were combined, vortexed, and aliquoted (500 µl per tube) for storage at -80°C.

For EB lysate preparation: cultures were transferred into 50 ml conicals and spun at 1500 rpm for five minutes at 5°C. Pellets were pooled and washed with cold PBS (three times with 50 ml each), spinning as previous between washes. EBs were lysed in NP40LB+PIs by sonication (three times, ten seconds each), aliquoted (500 µl/tube), and stored at -80°C. For other suspension cells (K562, Jurkat): cultures were transferred into 50 ml conicals and spun at 1500 rpm for five minutes at 5°C. Pellets were pooled and washed with cold PBS (two times with 25 ml each), spinning as previous between washes. Cells were lysed by resuspension in NP40LB+PIs, aliquoted (500 µl/tube), and stored at -80°C. All lysates were diluted to working concentrations in gel application buffer (2% SDS; 4% glycerol; 40 mM Tris-HCl, pH 6.8; 0.002% bromophenol blue; 1% β-mercaptoethanol) and boiled for three minutes.

Immunoblotting with commercial antibodies

Five µg/lane of each lysate was electrophoresed on seven cm 10% SDS-PAGE gels for 170 Vh, then transferred onto a nitrocellulose membrane at 100 volts for one hour. Membranes were stained with Ponceau (0.3% w/v in 3% trichloroacetic acid) and washed with distilled H_2O to identify molecular weight markers, followed by blocking in Tris buffered saline/Tween 20 (TBST; ten mM Tris-HCl, pH 7.4; 150 mM NaCl; 0.05% (v/v) Tween 20) with 5% (w/v) powdered milk for one hour at RT. Membranes were then incubated in TBST/5% milk containing the appropriate primary antibody (at an antibody-dependent dilutions) overnight at 4°C; washed five times for five minutes each with TBST/5% milk; incubated with a 1:10,000 dilution of horseradish peroxidase (HRP)-conjugated secondary antibody (species specificity dependent on primary

antibody) in TBST/5% milk for one hour at RT; washed five times for five minutes each in cold TBST buffer; and developed using an ECL chemiluminescent reagent (Pierce).

Primary antibodies and dilutions used include: rabbit polyclonal anti-Oct 4 antibody (Abcam, ab19857) at 0.2 μg/ml; rabbit monoclonal anti-Nanog antibody (Cell Signaling Technologies, D73G4) at 1:2000; mouse monoclonal anti-Tubulin beta-3 antibody (Millipore, 05-559) at 1:1000; rabbit polyclonal anti-NCAD antibody (Abcam, ab12221) at 1:1000; mouse monoclonal anti-CD31 antibody (DAKO, M0823) at 1:500; mouse monoclonal anti-beta actin antibody (Sigma, A2228, clone AC-74) at 1:20,000; rabbit polyclonal anti-B23 antibody (Covance) at 1:2000; rabbit polyclonal anti-cytokeratin 8 antibody (Abcam, ab52949) at 1:10,000; rat monoclonal anti-Hsc70 antibody (Stressgen, SPA-815) at 1:200,000; goat polyclonal anti-fetuin A antibody (Abcam, ab34505) at 1:5,000; mouse monoclonal anti-vinculin antibody (Sigma) at $1:10^6$.

Immunoblotting with patient sera

Five μg/lane of each lysate was electrophoresed on seven cm 10% SDS-PAGE gels for 170 Vh, then transferred onto a nitrocellulose membrane at 100 volts for one hour. Membranes were stained with Ponceau (0.3% w/v in 3% trichloroacetic acid) and washed with distilled H_2O to identify molecular weight markers, followed by blocking for one hour at room temperature (RT) in 3% bovine serum albumin in wash buffer (WB) (ten mM Tris-HCl, pH 7.6; 0.5% NP40; 150 mM NaCl). Membranes were then incubated with patient serum diluted in WB (1:5000) for one hour at RT. After extensive washing with WB (five times for five minutes each), membranes were subsequently

incubated with a horseradish peroxidase-conjugated goat anti-human IgG secondary antibody (Jackson Immunoresearch) at 1:10,000 dilution in WB for one hour. Membranes were then washed as above and incubated with enhanced chemiluminescent reagent (Pierce) followed by exposure to X-ray film.

2-Dimensional (2D) electrophoresis

For 2D electrophoresis, lysates were prepared in one of two ways: (*i*) cells from culture were washed twice with ten ml each cold PBS followed by washing twice with ten ml each cold [ten mM Tris HCl pH 7.4/205 mM Sucrose] and lysed directly into isoelectric focusing (IEF) buffer [eight M urea, two M thiourea, 4% CHAPS, 1% dithiothreitol (DTT) and protease inhibitors]; or (*ii*) lysates previously made in NP40LB were precipitated by TCA/acetone [briefly, a 1:1 volume of cold 20% trichloroacetic acid (TCA) was added to the lysate, followed by a 30 minute incubation on ice, pelleting at max speed 4°C for 15 minutes, supernatant removed, pellet washed with one ml ice cold acetone, repeated spin at max speed 4°C for 15 minutes, supernatant removed and pellet dried at RT] and resuspended in IEF buffer. First dimension focusing was performed using Bio-Rad ReadyStrip immobilized pH gradient (IPG) Strips and Protean IEF Cell per manufacturer's instructions. Seven cm IEF strips were used to determine the optimum pH range for the protein of interest while 17 cm IEF strips were used for sequencing. Briefly, each seven cm IEF strip was loaded with 130 µl consisting of 100 µg lysate, 1.5% 2-hydroxyethyl disulfide, 0.1% bromophenol blue and 0.4% ampholytes (pH dependent on IEF strip pH range). For 17 cm strips, 300 µg lysate was used in a total volume of 300 µl. Up to 12 IEF strips can be run simultaneously; these can

subsequently be processed immediately for the second dimension gel (see below), or they can be stored at -80°C. Prior to running the second dimension gels, IEF strips were incubated for 20 minutes at RT in ten mg/ml dithiothreitol, followed by 20 minutes in 25 mg/ml iodoacetamide, both in equilibration buffer (six M urea; 0.375 M Tris-HCl, pH 8.8; 2% SDS; 20% glycerol). IEF strips were then loaded onto 10% SDS PAGE gels (16 cm or seven cm gels, depending on the size of the IPG strips), covered with 1% low melt agarose (Promega) and electrophoresed as described above (16 cm gels were electrophoresed for 850 volt hours, and transferred onto nitrocellulose membranes for four hours at 400 mA). To ensure consistency, IEF strips run in parallel were used for 2D immunoblotting and corresponding protein sequencing experiments.

2D immunoblotting and antigen localization

2D gels were transferred to nitrocellulose membranes, stained with Ponceau to visualize resolved protein spots, and scanned at high resolution for alignment with the subsequent immunoblots. Immunoblots were performed as described. During X-ray film exposure, precise documentation of the membrane perimeter on the film is critical for subsequent alignment of the exposed film with a printout of the Ponceau scan. This enables the immunoblotted spot of interest to be aligned with the corresponding Ponceau-stained protein spot. Once the protein of interest is located on the Ponceau scan, nearby spot patterns can be used for orientation to find the same spot on the preparative gel for plucking and protein identification.

Protein identification

Preparative SDS-PAGE second dimension gels were made using solutions and procedures that minimize exposure to airborne contaminants. All glassware and equipment were only used for proteomics experiments and were washed twice with 0.1% SDS (w/v) followed by extensive rinsing with ultrapure distilled water and 100% methanol to minimize protein contamination. After performing first and second dimension gel runs as described, gels were fixed in destain (14% v/v acetic acid; 24% v/v methanol) for 15 minutes, incubated for one hour in Gel Code Blue Stain (Thermo Scientific) and then destained in ultrapure distilled water for two to four hours. Destained gels were placed on a light box to facilitate visualization of stained protein spots, and gel plugs containing the protein spot of interest and a nearby "blank" gel plug containing no visible protein were carefully removed from the preparative gel using a sterile pipette tip. The gel plugs were immediately washed twice in 50% methanol for ten minutes each with shaking and stored at -20°C. The remaining preparative gel was washed three times for five minutes each in running buffer (25 mM Tris-HCl; 200 mM glycine; 0.1% (w/v) SDS), and proteins were transferred to nitrocellulose membrane for immunoblotting to verify successful plucking. Gel plugs confirmed to contain the protein of interest were sent to the Johns Hopkins Mass Spectrometry and Proteomics Facility for liquid chromatography tandem mass spectrometry (LC/MS/MS) peptide sequencing.

Antigen identity confirmation

Solubility fractioning – 50 μg of cell lysates were brought up to 90 μl in NP40LB/PIs and solubilized by sonication twice for ten seconds each (amplitude 3-4) followed by vortexing. The insoluble lysate fractions were collected by centrifugation at

4°C for 30 minutes at max speed (13.2 rpm) in tabletop centrifuge and supernatants containing soluble fractions were moved to a new tube. Ten µl of 5x GAB/βMe was added to each soluble fraction. Insoluble fractions were resuspended in 90 µl NP40LB/PIs by sonication twice for ten seconds each (amplitude 3-4) followed by vortexing and ten µl of 5x GAB/βMe was added. Samples were boiled three minutes and 11 µl was loaded per lane for immunoblotting.

2D membrane stripping – Following immunoblot with commercial (non-human) primary antibody, membrane was washed five minutes with TBST then stripped in 0.1 M glycine, pH 2.7, for seven minutes at 37°C following by shaking at RT for seven minutes. Membrane was then washed twice for five minutes each with WB/azide, then washed three times for five minutes each with WB(-) azide. To confirm stripping, membrane was incubated in ECL reagent for five minutes and used to expose X-ray film for five to ten minutes. After antibody stripping was confirmed, membrane was washed for five minutes in WB(-) azide, blocked 30 minutes in WB/azide/3% BSA and blotted with human serum as previously described.

Overlay program – Digital overlay of 2D immunoblots was performed using a computer program designed specifically for this purpose by Steven J. Handy (Senior Software Engineer, The Johns Hopkins University). This program allows the import of scanned X-ray films, which can then be modified (scaling, transparency, and color tinting) to allow digital overlay of immunoblot images. Precise alignment of images was possible through careful recording of the edges of the nitrocellulose membrane on the X-ray film as well as the positions of the molecular weight markers.

Transfection – U2OS and HEK293 cells were transfected with pKeratin 8 (Origene) and pFetuin A (Origene) respectively using FuGene 6 (Roche) per manufacturer's instructions. Mock transfected cells were used as negative controls.

Non-reduced lysate preparation – For preparation of non-reduced gel samples, lysates were diluted to working concentrations in gel application buffer (2% SDS; 4% glycerol; 40 mM Tris-HCl, pH 6.8; 0.002% bromophenol blue) without β-mercaptoethanol and boiled for three minutes. Reduced and non-reduced EB lysates were resolved on a 7.5% SDS-PAGE gel and immunoblotted as previously described.

Deglycosylation – Deglycosylation of EB lysates was performed using a Protein Deglycosylation Mix (New England Biolabs, P60395) containing PNGase F, O-Glycosidase, Neuraminidase, β1-4 Galactosidase, and β-N-acetylglucosaminidase per manufacturer's instructions.

EB embedding

All steps were performed gently to preserve morphology. Embryoid bodies from culture were pelleted at 500 rpm for two minutes at RT. The media was removed and the pellet was washed in five ml cold PBS by inversion. The EBs were spun as previous, PBS removed, and pellet fixed in five ml 2% PLP (2% paraformaldehyde, 0.075 M L-Lysine, 0.0375 M PO_4, 0.01 M Na-*m-per*iodate) for a total of 30 minutes at RT with occasional inversion. PLP was removed following a spin as previous and fixation was quenched by addition of five ml fresh NH_4Cl (50 mM NH_4Cl in PBS, made right before use), followed by inversion. The NH_4Cl was removed following a spin as previous and ~20 µl 2% low melt agarose (in H_2O +colorant, i.e. ferritin; @65°C) was added and

allowed to partially solidify at room temperature for two minutes. For cryoprotection, ten ml cold [0.5 M sucrose in 0.1 M $NaPO_4$, pH 7.4] was added and complete solidification was performed at 4°C for four to eighteen hours. The agarose plug was transferred to a cryomold containing TBS freezing medium and covered fully with additional freezing medium followed by immersion into 2-methylbutane cooled with dry ice. After freezing for two to five minutes the cryomold was dried on dry ice followed by storage at -80°C. Ten µm sections were cut using a cryostat at -20°C and sections were placed on lysine coated glass microscopy slides (stored at -20°C).

Immunofluorescence

Tissue sections were encircled with a hydrophobic pen, then fixed and permeabilized by dipping slides into ice cold acetone for 30 seconds and immediately washed by immersion into PBS three times 30 seconds (without letting slides dry). Sections were blocked for one 1 hour at RT in 5% serum* in PBS (*using serum from same animal as the secondary antibody). Primary antibodies were diluted in a final volume of 250 µl per slide (dilutions detailed below) in 5% serum* in PBS and spun at 5000 rpm for five minutes before use. Slides were incubated in primary antibody dilutions overnight at 4°C in a humidity chamber, then washed in PBS three times for ten minutes each without shaking. Secondary antibodies were diluted 1:200 in 1% normal goat serum in PBS and spun at 5000 rpm for five minutes prior to adding 100 µl per slide and incubated for one hour at RT in the dark. Slides were then washed in PBS three times for ten minutes each and, while still wet, one to two drops of Prolong Gold w/ DAPI (Life Technologies, P-36931) was added to each tissue section, covered with a

glass cover slip, allowed to set overnight at RT with minimal light exposure and store in the dark at 4°C.

Primary antibodies and dilutions used include: rabbit polyclonal anti-Oct 4 antibody (Abcam, ab19857) at 1:400; rabbit polyclonal anti-NCAD antibody (Abcam, ab12221) at 1:300; mouse monoclonal anti-Tubulin beta-3 antibody (Millipore, 05-559) at 1:300; anti-RNA polymerase III positive scleroderma patient serum at 1:1000; anti-topoisomerase 1 positive scleroderma patient serum at 1:1000; mouse monoclonal anti-CD31 antibody (DAKO, M0823) at 1:20. Corresponding secondary antibodies specific for primary antibody species and conjugated with either Alexa Fluor 488 or Alexa Fluor 594 (Life Technologies) were used as appropriate at a 1:200 dilution.

Results

Detection of novel scleroderma autoantigens in differentiating stem cells

Undifferentiated human embryonic stem cells were allowed to differentiate spontaneously into three-dimensional embryoid bodies, which were harvested at various time points during differentiation. With continued differentiation, the embryoid bodies increase in size and become increasingly cystic, with representative light microscopy images of an undifferentiated ES cell colony, and days three, seven, and ten of EB differentiation shown (Figure 9A-D respectively). Differentiation is associated with down-regulation of the pluripotency markers Oct 4 and Nanog (Figure 9E), as well as up-regulation of germ layer and lineage specific markers, including Tubulin beta-3 (ectoderm), N-cadherin (NCAD) (mesoendoderm), and CD 31 (endothelial lineage) (Figure 9E). Expression of scleroderma autoantigens (topo 1, snRNP70/U1-70K, and B23) were also evaluated in differentiating embryoid bodies. Topo 1 and snRNP70 followed similar expression patterns, with high expression in undifferentiated ES cells and low expression in differentiation days one through four, followed by increasing expression that plateaus around differentiation day ten (Figure 9E). In contrast, B23 expression levels are highest in undifferentiated and day one EBs, but also transiently drop around differentiation day four (Figure 9E). In addition to undifferentiated ES cells, day ten EBs were chosen for subsequent screens based on increased expression of scleroderma autoantigens and the presence of a mixture of cell lineages, including cells expressing CD31, consistent with reports that EBs at this stage are enriched for vascular progenitor cells *(68)*.

Scleroderma patient sera were screened by immunoblotting for high titer autoantibodies targeting proteins expressed in undifferentiated human embryonic stem cells or day ten human embryoid bodies, with HeLa cells as a control for the detection of ubiquitously expressed proteins (Figure 10A). Over one third of 89 randomly selected scleroderma sera tested contained autoantibodies targeting proteins exclusively present in undifferentiated ES or day ten EB lysates (Table 3). As a more stringent test of our hypothesis, we next screened the sera of 35 patients clinically determined to be ANA negative. Strikingly, over half had autoantibodies targeting proteins expressed in undifferentiated ES cells or day ten EBs that are not expressed by HeLa cells (Figure 10B). Finally, 23 patients with scleroderma and cancer were screened and many were found to have autoantibodies targeting ES or EB specific proteins (data not shown, Table 3). In contrast, no high titer autoantibodies were detected in 17 healthy control sera tested, a representative example is denoted (HC) in Figure 10A. Patient sera were next used as probes in the identification of the proteins targeted by autoantibodies in patients previously considered autoantibody negative.

Identification of keratin 8 as a novel scleroderma autoantigen

The identification of the ~50 kDa protein targeted by serum autoantibodies in patient SSc-3 (Figure 10B) using the previously described triangulation strategy is illustrated (Figure 11). This antigen was of particular interest because it is targeted by a patient previously determined to be negative for antinuclear autoantibodies and the targeted protein was not expressed by several commonly used cell lines, including HeLa (cervical cancer derived), Jurkat (T lymphocyte derived), U2OS (osteosarcoma derived),

and K562 (erythroleukemia derived) (Figure 11A). The protein was localized on a 2D gel (pH 3-10) by immunoblotting (Figure 11B) and the location of the protein was noted on the corresponding Ponceau stain (data not shown). An identical gel was stained with colloidal blue and multiple spots containing the protein of interested were plucked from the gel (Figure 11B & 11C insets) and confirmed to contain the protein of interest by post-pluck immunoblotting with SSc-3 serum (Figure 11C). Three of the plucked gel plugs were sent for separate LC/MS/MS peptide sequencing.

After subtracting contaminants sequenced from a nearby "blank" gel plug, the only protein detected in all three gel plugs was keratin 8 (KRT8). The significance of this finding was initially unclear because keratins are the most common contaminants detected by LC/MS/MS peptide sequencing *(69)*. To determine if the SSc-3 targeted antigen could be a keratin, day ten EB lysates were separated into soluble and insoluble fractions and immunoblotted with SSc-3 serum (Figure 11 D). Whereas vinculin was localized to the soluble fraction as expected, the SSc-3 targeted protein localized to the insoluble fraction, as expected for a keratin.

Confirmation of the identity of this protein was pursued using multiple methods. First, EB lysate resolved by 2D gel (pH 3-10) was immunoblotted with a rabbit polyclonal anti-cytokeratin 8 antibody (Figure 12A). The membrane was subsequently stripped of all bound antibody (Figure 12B) and re-probed with the SSc-3 human serum (Figure 12C). The alignment of the two blots confirms precise overlap of the proteins bound by the polyclonal anti-keratin 8 antibody and SSc-3 patient serum autoantibodies (Figure 12D). A second method of confirmation was attempted using ^{35}S-methionine labeled keratin 8 (^{35}S-KRT8) generated by IVTT using keratin 8 cDNA. While the anti-

keratin 8 commercial antibody was able to immunoprecipitate the recombinant ^{35}S-KRT8, the SSc-3 patient serum was not (data not shown). To assess SSc-3 antibody recognition of keratin 8 in a more physiological state, a plasmid expressing keratin 8 cDNA (pKeratin 8) was transiently transfected into U2OS osteosarcoma cells (Figure 12E). Immunoblotting with the polyclonal anti-cytokeratin 8 antibody demonstrated the absence of keratin 8 in mock transfected cells (-) and robust expression of keratin 8 in the transfected cells (+). The SSc-3 patient serum also recognized the same 50 kDA only in the transfected cells, in contrast to no detection by a healthy control serum. Lysates from U2OS cells transiently expressing keratin 8 were separated into soluble and insoluble fractions and used to screen 63 additional scleroderma sera for the presence of anti-keratin 8 antibodies by immunoblotting (data not shown). Only 1/63 scleroderma patients was positive for anti-keratin 8 antibodies by immunoblotting, compared to 0/20 healthy controls. Both the prototype and screen anti-keratin 8 positive patients have the limited form of scleroderma with mild disease.

Identification of fetuin A as a novel scleroderma autoantigen

The identification of the antigens targeted by autoantibodies in patient SSc-4 was pursued because this patient had an intriguing specificity pattern (Figure 10B). Two similarly sized proteins are targeted, the larger of which is ~70 kDa and expressed equally in all cells tested (Figure 10B open arrow) and the other is ~68 kDa and faintly detected in HeLa cells, absent in undifferentiated ES cells, and strongly upregulated in day ten EBs (Figure 10B, black arrow). This patient also has a history of cancer and is clinically negative for known SSc autoantibodies.

Although these two proteins were difficult to separate by size alone on a 1D gel, they demonstrated substantially different isoelectric points (pIs) when separated on a 2D gel with a pH 3-10 gradient (Figure 13A), with the ~68 kDa EB specific antigen having a pI near three (black arrow, lower left inset) and the ~70 kDa ubiquitously expressed protein having a pI near 5.5 (open arrow, lower right inset). In addition, both of these proteins demonstrate multiple horizontally separated isoforms on the 2D gel, likely due to post translational modifications that alter the pI without a significant change in protein size. The protein spots corresponding to the immunoblot signal were identified and plucked from an identically run gel code blue stained gel (Figure 13B) and correct spot selection was confirmed (Figure 13C) as previously described.

The ~70 kDa commonly expressed protein band was identified by LC/MS/MS sequencing to be heat shock cognate 71 kDA protein (HSC70), which was confirmed by precise overlay of identically run 2D gels immunoblotted with a commercial anti-HSC70 antibody and SSc-4 patient serum (Figure 13D).

The leading candidate for the acidic 68 kDa antigen expressed in day ten EBs after LC/MS/MS sequencing was fetuin A, a protein normally synthesized in the liver and secreted into the plasma. This protein undergoes multiple post-translational modifications prior to secretion, including cleavage into two peptide chains (of 282 aa and 27 aa) held together by a disulfide bond and extensive glycosylation of the larger chain A. Based on the size, pI, and 2D pattern of the antigen of interest (and that lysates were regularly reduced prior to resolution), it is likely that the 68 kDa band represents the glycosylated form of the 282 aa A chain of fetuin A, predicted to be only 30 kDa in the absence of glycosylation.

Several approaches were taken to confirm that the 68 kDa SSc-4 antigen was fetuin A. First, reduced and non-reduced lysates were probed with a polyclonal anti-fetuin A antibody and SSc-4 serum (Figure 13E). While the 68 kDa protein (black arrow in SSc-4 blot) is present in the reduced lysates (+) as previously shown, the protein is absent in the non-reduced lysates (-). As expected, in the SSc-4 immunoblot, the protein band corresponding to HSC70 (Figure 13E, open arrow) remains unchanged in the absence of reduction.

Deglycosylation of the EB lysates prior to immunoblotting also confirmed a parallel decrease in band size in the deglycosylated lysates (Figure 13F (+)) with the polyclonal anti-fetuin A antibody (upper panel) and the SSc-4 serum (lower pane, black arrows). Notably, deglycosylation did not impact the ability of the patient autoantibodies to bind fetuin A. Immunoprecipitation of ^{35}S-methionine labeled fetuin A (^{35}S-FETA) generated by IVTT using fetuin A cDNA was also attempted. Although the polyclonal anti-fetuin A antibody was able to IP the ^{35}S-FETA, the SSc-4 patient serum was not (data not shown). Finally, a plasmid containing fetuin A cDNA (pFetuin A) was transiently transfected into the HEK293 (human embryonic kidney derived) cell line. Both the commercial anti-fetuin A antibody and SSc-4 bound the 68 kDa protein exclusively in the transfected (+) cells, with no protein of this size detected in the mock transfected control (-) (Figure 13G).

To determine the prevalence and possible clinical relevance of the newly discovered anti-fetuin A autoantibodies, 63 randomly selected scleroderma patient sera were screened by immunoblotting lysates from HEK293 cells transfected with either a plasmid containing fetuin A cDNA or a mock transfected control (data not shown). Six

patients were positive for anti-fetuin A autoantibodies by immunoblotting (9.5%), with 5/6 having the diffuse form of scleroderma. In contrast 0/20 healthy controls had anti-fetuin A autoantibodies.

Localization of autoantigen expression by indirect immunofluorescence

Finally, to determine which cell lineages within the heterogeneous embryoid bodies are expressing high levels of scleroderma autoantigens, EBs at various stages of differentiating were carefully embedded to preserve their 3-dimensional structure and antigen expression was detected using indirect immunofluorescence (Figure 14). By day seven of differentiation, the EBs demonstrate a consistent expression pattern of germ layer markers, with undifferentiated (Oct 4 positive) cells remaining at the periphery (Figure 14A), and mesoendodermal (NCAD positive) cells and ectodermal (Tubulin beta-3 positive) cells in the center of the EB (Figures 14B and 14C respectively). Colocalization of RNA polymerase III expression with Oct 4 revealed that RNA polymerase III is upregulated in cells expressing Oct 4 in day seven EBs (Figure 14D). In contrast, topoisomerase 1 expression levels appeared relatively uniform in day seven EBs and are not upregulated in Oct 4 expressing cells (Figure 14E). Finally, since we hypothesized that scleroderma autoantigens may be upregulated in blood vessel progenitors, we sought to colocalize expression of RNA polymerase III and topoisomerase 1 with expression of the early vessel marker CD31 in day 12 EBs. The nuclear staining of RNA pol III (Figure 12F, red) was not upregulated in the cells expressing the membrane protein CD 31 (Figure 12F, green). However, it does appear that CD31 negative cells near the site of blood vessel formation show higher expression

of RNA polymerase III than surrounding cells (Figure 14F, arrows). In contrast, there is no apparent association of topoisomerase 1 expression (Figure 14G, red) with nascent vessels expressing CD31 (Figure 14G, green) in day 12 EBs. These data demonstrate differential expression of the scleroderma autoantigen RNA polymerase III in differentiating cells within embryoid bodies, with increased expression seen in undifferentiated or early ectodermal Oct 4 positive cells in day seven EBs and in CD31 negative cells near sites of blood vessel formation in day 12 EBs.

Discussion

This work has shown that roughly one third of randomly selected scleroderma sera and a striking 60% of ANA negative scleroderma sera contain autoantibodies that recognize antigens uniquely expressed in undifferentiated stem cells or differentiating day ten embryoid bodies. Moreover, none of these proteins are targeted by healthy control sera and they are not expressed in HeLa cells, which are classically used for antigen identification. These data indicate that many antigens important in scleroderma autoimmunity remain to be identified. Moreover, the discovery of many proteins targeted by high titer autoantibodies in ANA negative patient sera confirm that autoimmunity is a prominent aspect of their disease. Finally, the abundance of novel antigens detected specifically in ES or EB cells strongly supports the hypothesis that stem and progenitor cells are improved approximations for the true target(s) of the autoimmune response in scleroderma.

Keratin 8 was identified as a novel autoantigen in scleroderma. Current knowledge of the function of keratin 8 suggests many intriguing mechanistic possibilities for the role of anti-keratin 8 autoantibodies in scleroderma. For example, keratin 8 is a clinically important marker of epithelial cancers *(70, 71)*. Moreover, the C terminus of keratin 8 penetrates the cell membrane and has been identified as a major plasminogen receptor on the surface of breast cancer cells, facilitating plasmin activation and the degradation of basement membranes and extracellular matrix components. Interestingly, plasminogen binding and activation can be inhibited by the presence of a monoclonal antibody generated against the C-terminus of keratin 8 *(72)*. These data suggest that the generation of anti-keratin 8 autoantibodies would be potentially effective component of

an anti-tumor immune response. Evaluation of patient anti-keratin 8 autoantibodies for C-terminal epitope specificity and the ability to block fibrin activation by cancer cells would support the model that these autoantibodies are generated during an anti-tumor immune response.

In addition to its role in cancer, keratin 8 has also been implicated in the injury response of vascular smooth muscle cells *(73)* and epithelial cell resistance to stress and apoptosis *(74, 75)*. A role in the response to tissue injury may lead to high levels of keratin 8 expression at sites of active disease in scleroderma. Moreover, the presence of anti-keratin 8 antibodies with the capacity to block the activation of fibrinogen in the context of tissue repair could result in a profibrotic environment like that seen in scleroderma. Comparing the levels of keratin 8 expression in diseased and healthy tissues may lead to important insights regarding the role of this autoantigen in scleroderma pathogenesis.

More extensive screening for the presence of anti-keratin 8 antibodies will be necessary to determine their prevalence, specificity, and any clinical phenotype associations. Interestingly, the patient anti-keratin 8 autoantibodies were not able to immunoprecipitate a recombinantly expressed form of the protein, which may suggest that post-translational modifications may be important in recognition of this antigen. In fact, keratin 8 undergoes extensive phosphorylation, acetylation *(76, 77)*, and O-linked N-acetylglucosamine modification *(78)*, which could be relevant to a unique function in progenitor or cancer cells as well as antibody recognition. Whereas radiolabeling cells transfected with keratin 8 may allow more extensive screening by immunoprecipitation, a

method that is more sensitive than immunoblotting, it may also be interesting to explore the potential modification dependent recognition of this protein by patient autoantibodies.

The identification of fetuin A (also known as α_2-HS glycoprotein) as a target of scleroderma patient autoantibodies could be another exciting discovery. This protein is secreted by the liver into the plasma, is found at much higher levels in fetal serum than adult serum, and is a negative acute phase reactant *(79)*. Fetuin A is also a major component of fetal bovine serum and has been implicated as the key molecule facilitating cell attachment and growth in tissue culture *(80, 81)*.

Fetuin A has also been implicated as a tumor antigen. In a mouse model of lymphoma, comparison of serum antibodies found that mice immunized with a tumorigenic lymphoma line developed anti-fetuin A antibodies whereas mice immunized with a non-tumorigenic variant line did not, despite the fact that fetuin A expression levels are similar in both cell lines. These data suggest that a modification of fetuin A associated with tumorigenicity is also capable of specifically initiating an anti-fetuin A immune response *(82)*. In a mouse model of breast cancer, the absence of fetuin A decreases the incidence of breast cancer and increases tumor latency *(83)*, while fetuin A was also found to promote the attachment and growth of human breast cancer cells *in vitro (80)*. Moreover, anti-fetuin A antibodies have been suggested as a potential biomarker in breast cancer screening *(84)*. Given the temporal relationship between the onset of breast cancer and scleroderma in some patients *(85)*, these data support a role for fetuin A as a tumor antigen that may be targeted by autoantibodies in scleroderma. Examining expression levels and post-translational modifications of fetuin A in

scleroderma disease tissues and scleroderma associated tumors may provide insight into the pathogenic mechanisms underlying this disease.

Interestingly, a fetuin A knockout mouse model demonstrates an important role for this protein in the inhibition of ectopic calcification *(86)* and there is evidence that direct binding of calcium and phosphate to an array of acidic amino acid residues in one domain of the protein prevents basic calcium phosphate precipitation *(87)*. This is of particular relevance to scleroderma because subcutaneous calcium deposition, known as calcinosis, can be a prominent feature of disease. If the presence of anti-fetuin A autoantibodies inhibits the ability of fetuin A to bind calcium, this could be the first hint at a mechanism for calcinosis. Looking at the correlation of calcinosis and anti-fetuin A autoantibodies among scleroderma patients will be useful in determining whether or not these antibodies play a pathogenic role.

Given the many potential roles that fetuin A may play in the pathogenesis of scleroderma, we are eager to explore the prevalence and clinical associations of anti-fetuin A antibodies in our scleroderma cohort. Initial experiments indicate that the serum antibodies are not able to immunoprecipitate the IVTT generated protein, indicating that post-translational modifications may be necessary for recognition. Immunoblotting or immunoprecipitation of transfected lysates will potentially allow a screen of many patients and normal controls to determine the significance of this autoantibody in scleroderma.

We demonstrate that expression of known scleroderma autoantigens may be differentially regulated during stem cell differentiation. Whereas most scleroderma autoantigens identified to date perform necessary cellular functions and are thus

ubiquitously expressed, characterization of these proteins during stem cell differentiation has not been studied to our knowledge. Characterization of autoantigen levels during differentiation of human embryonic stem cells is challenging because not all scleroderma autoantigens are amenable to immunoblotting (including RNA polymerase III and centromere A proteins). In addition, EBs are heterogeneous mixtures of cell types at various stages of differentiation, such that upregulation of antigen expression by a given cell type within such a mixture may not be consistently detected by immunoblots of EB lysates. Detection of antigen expression on an individual cell level by indirect immunofluorescence may be useful in identifying relatively higher levels of antigen expression in particular lineages within these EBs, as our preliminary data suggest. Additional studies colocalizing autoantigen expression and cell lineage markers may provide critical insights regarding the true target of the immune response in scleroderma. These studies will also be greatly enhanced by characterization of cells expressing high levels of autoantigens in the target tissues from scleroderma patients – studies that are currently being pursued by the Rosen lab.

This investigation is the first to demonstrate the targeting of progenitor specific proteins by autoantibodies in scleroderma, including the discovery of two novel autoantigens, keratin 8 and fetuin A. Our data also suggest that well established targets in scleroderma, including RNA polymerase III, are differentially regulated during stem cell differentiation. Together, these data support the hypothesis that differentiating progenitors involved in tissue repair may be a target of the immune response in scleroderma.

Table 3: Immunoblot screen for antibodies targeting proteins expressed by undifferentiated human embryonic stem cells and differentiating embryoid bodies.

Serum Screened	Total (n)	Positive[a] n (%)
Scleroderma	89	32 (36%)
ANA Negative SSc	35	21 (60%)
SSc-Cancer	23	8 (35%)
Total	147	61 (41%)
Healthy Controls	17	0 (0%)

[a] Positive for antibodies binding ≥ 1 protein in ES and/or EB lysates in the absence of a corresponding protein in HeLa lysate

SSc = Scleroderma; ANA = anti-nuclear antibodies; ES = undifferentiated human embryonic stem cells; EB = day ten differentiating embryoid bodies.

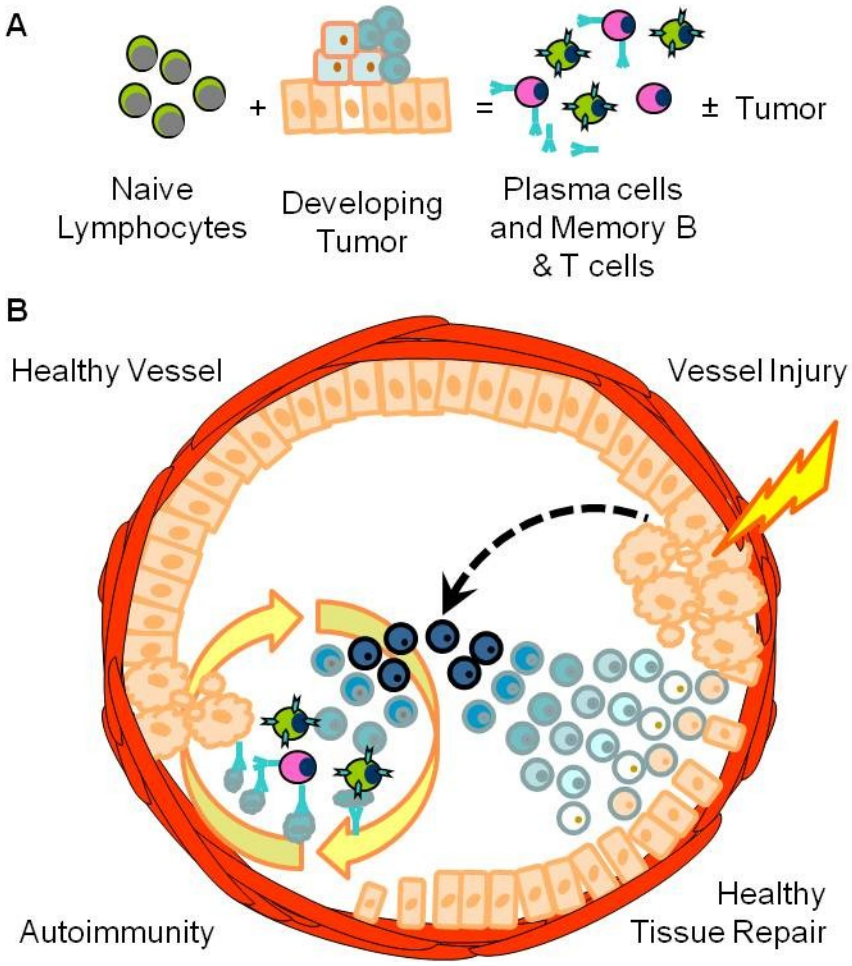


Figure 8: A model for the mechanism of autoimmunity in scleroderma. A. The immune system combats a nascent tumor resulting in mature lymphocytes specific for self-antigens expressed by the neoplasm, which may include proteins that are normally only expressed by stem/progenitor cells. B. Tissue injury results in the expansion and differentiation of stem and progenitor cells involved in the tissue repair process. In the presence of auto-reactive lymphocytes, progenitors expressing previously sensitized tumor antigens are targeted for destruction by the immune system. The failure of healthy tissue repair leads to a feed forward cycle of immune mediated tissue damage and sustained stimulation of the immune system by antigen expressing progenitor cells attempting tissue repair.

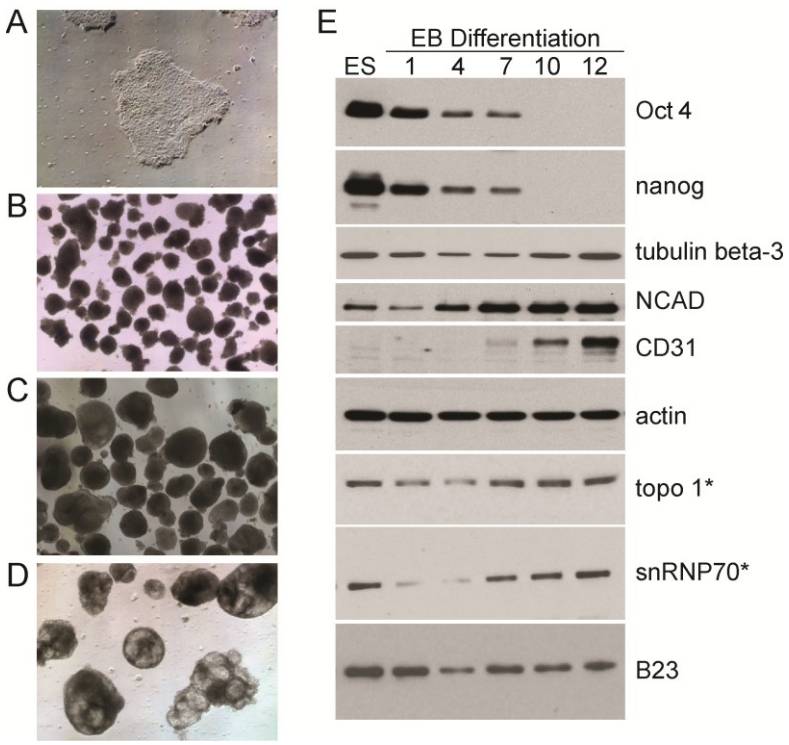


Figure 9: Differentiation of human embryonic stem cells into embryoid bodies. Light micrographs of adherent, undifferentiated ES colony (A), and day three (B), day seven (C), and day ten (D) EBs in suspension. There is significant heterogeneity in the size and shape of EBs throughout differentiation, with both size and cystic content increasing with prolonged differentiation. (E) Immunoblots of undifferentiated ES and differentiating EBs at days one, four, seven, ten, and 12 demonstrating down-regulation of pluripotency markers (Oct 4, Nanog) and upregulation of germ layer/lineage markers (Tubulin beta-3 (ectoderm), N-cadherin (NCAD) (mesoendoderm), and CD 31 (endothelial lineage)). Expression levels of known scleroderma autoantigens (topoisomerase 1, RNP, and B23) are lowest on differentiation day four and then increase, plateauing around day ten. Actin immunoblot demonstrates equal protein loading in each lane. *Topo 1 and snRNP70 immunoblots performed using scleroderma patient sera (all other proteins detected with commercial antibodies).

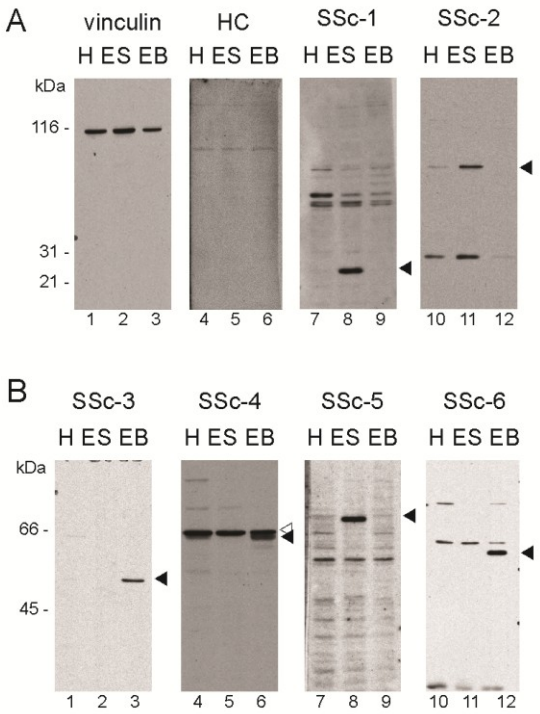


Figure 10: Patient sera detect novel scleroderma autoantigens in stem cells and differentiating progenitor cells. (A) Immunoblot screen of HeLa cell (H), undifferentiated embryonic stem cell (ES), and day ten differentiating embryoid body (EB) lysates with randomly selected scleroderma patient sera (SSc-#). Note that patients SSc-1 and SSc-2 have high titer autoantibodies to proteins expressed only in undifferentiated ES cells (lanes 8 & 11, black arrows). Vinculin immunoblot demonstrates protein loading in each lane and a representative healthy control serum (HC) demonstrates the absence of high titer autoantibodies. (B) Antinuclear antibody negative SSc sera contain autoantibodies targeting proteins exclusively expressed in ES cells (lane 8, black arrow) and differentiating EBs (lanes 3, 6, 12, black arrows), as well as occasional proteins expressed by all cell types tested (SSc-4, white arrow).

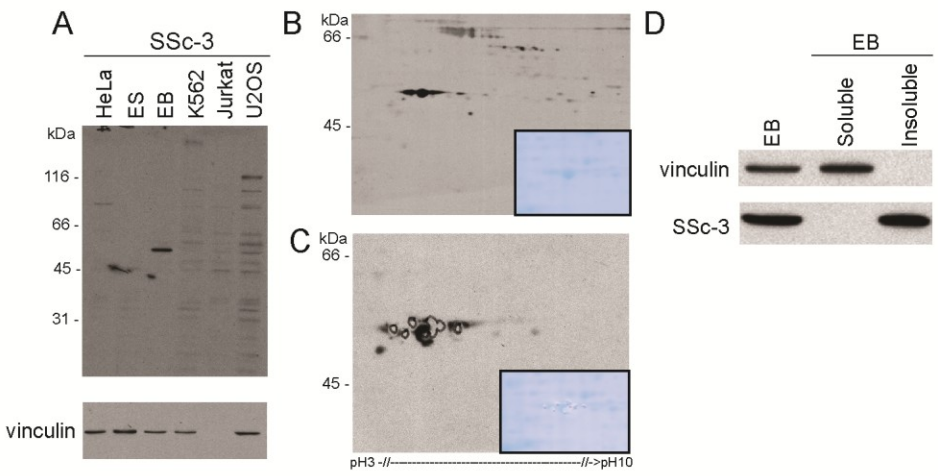


Figure 11: Identification of the autoantigen targeted by SSc-3. (A) Immunoblot of HeLa, ES, and EB lysates as well as additional cell lines (K562 (erythroleukemia), Jurkat (T lymphocyte), U2OS (osteosarcoma)) with SSc-3 scleroderma patient serum. The 50 kDa protein targeted by SSc-3 autoantibodies is only seen in the EB lysate. Vinculin immunoblot (below) shows protein loading in each lane (note: vinculin is not expressed by Jurkat cells). (B) SSc-3 immunoblot of EB lysate after 2D resolution. First dimension: pH 3-10 IEF; second dimension: 10% SDS-PAGE. (C) SSc-3 immunoblot of identically run gel after removal of gel plugs containing the protein of interest. Insets: corresponding colloidal blue stained gels before (B, inset) and after (C, inset) removal of gel plugs, prior to transfer to nitrocellulose membranes. (D) Immunoblot of EB lysates fractions showing the SSc-3 targeted protein is insoluble, while vinculin remains in the soluble lysate fraction.

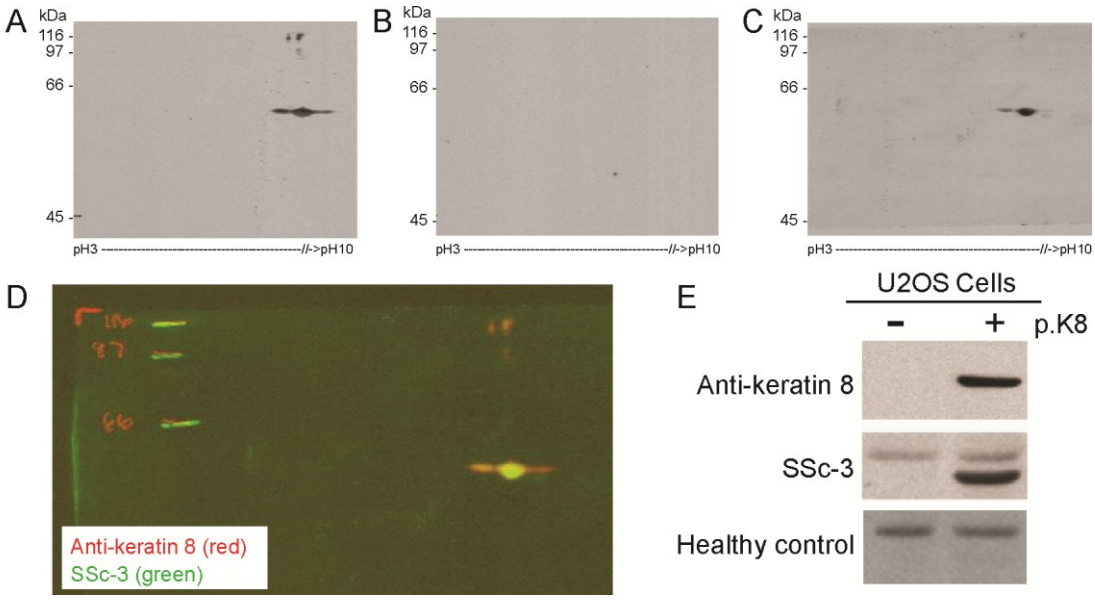


Figure 12: Confirmation of keratin 8 as a novel scleroderma autoantigen. (A) Immunoblot with rabbit polyclonal anti-keratin 8 antibody of EB lysate after 2D resolution. First dimension: pH 3-10 IEF; second dimension: 10% SDS-PAGE. (B) The nitrocellulose membrane was subsequently stripped of antibody and exposure after incubation with a chemiluminescent reagent confirms the absence of previously bound anti-rabbit HRP-conjugated secondary antibody. (C) The membrane was then re-probed with scleroderma patient serum SSc-3 and an HRP-conjugated anti-human secondary antibody. (D) Precise digital overlay of X-ray film images confirms exact alignment of signal from bound polyclonal anti-keratin 8 Ab (red) and scleroderma patient serum SSc-3 (green). (E) Immunoblots of mock (-) or pKeratin8 (+) transfected U2OS cells with a polyclonal anti-keratin 8 antibody, SSc-3 patient serum, and healthy control serum. The 50 kDa keratin 8 protein is only detected in the pKeratin 8 transfected cells by the polyclonal anti-keratin 8 antibody and the SSc-3 patient serum, but is not detected in the mock transfected cells or by immunoblot with healthy control serum. HRP = horseradish peroxidase.

Figure 13: Identification of HSC70 and fetuin A as novel scleroderma autoantigens. (A) Immunoblot with SSc-4 patient serum of EB lysate after 2D resolution. First dimension: pH 3-10 IEF; second dimension: 10% SDS-PAGE. Note separation of the acidic 68 kDa EB antigen (black arrow) and the ubiquitously expressed 70 kDa antigen (open arrow). (B) Identically run 2D gel stained with colloidal blue and with gel plugs removed for LC/MS/MS sequencing. (C) Following transfer to a nitrocellulose membrane, immunoblotting with SSc-4 patient serum confirms plucked gel plugs contain the antigens of interest. Areas of emphasis boxed and enlarged directly below. (D) 2D immunoblot with monoclonal anti-HSC70 antibody (left), identical gel blotted with SSc-4 patient serum (middle), and precise digital overlay of X-ray images (right; HSC70 in red, SSc-4 in green). (E-F) Immunoblots comparing protein bands detected with the SSc-4 patient serum to those detected with a polyclonal anti-fetuin A antibody: (E) Reduced (+) and non-reduced (-) EB lysates - note that the 68 kDa antigen (black arrow) is detectable exclusively in the reduced lysates; (F) EB lysates before (-) and after (+) deglycosylation – note the corresponding reduction in size of the 68 kDa bands with deglycosylation (black arrows); (G) HEK293 (human embryonic kidney) cells mock (-) and pFetuin A (+) transfected – note detection of the 68 kDa band exclusively in the pFetuin A transfected cells by both the commercial anti-fetuin A antibody and the SSc-4 patient serum. For SSc-4 blots, HSC70 is indicated by white arrows and the 68 kDa EB antigen is indicated with black arrows.

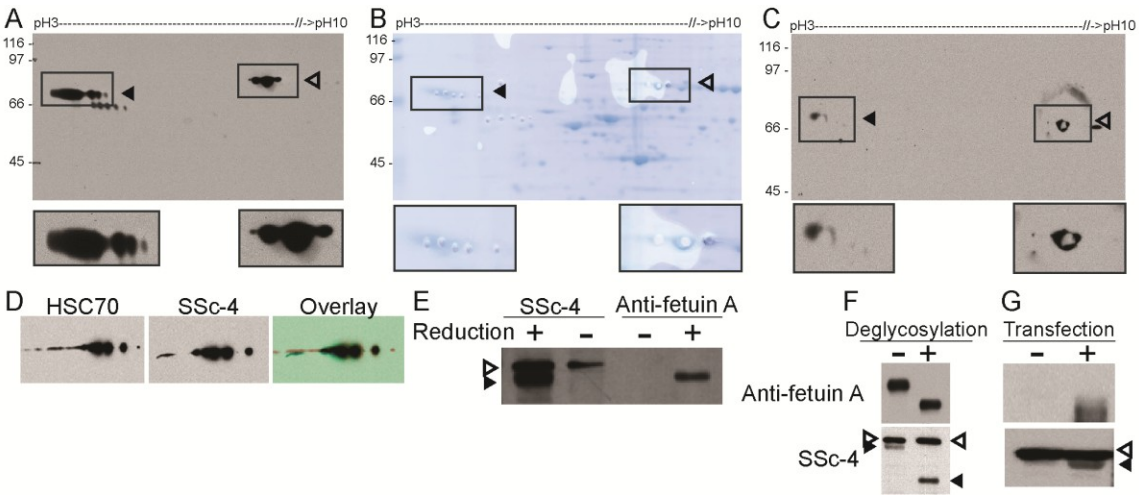
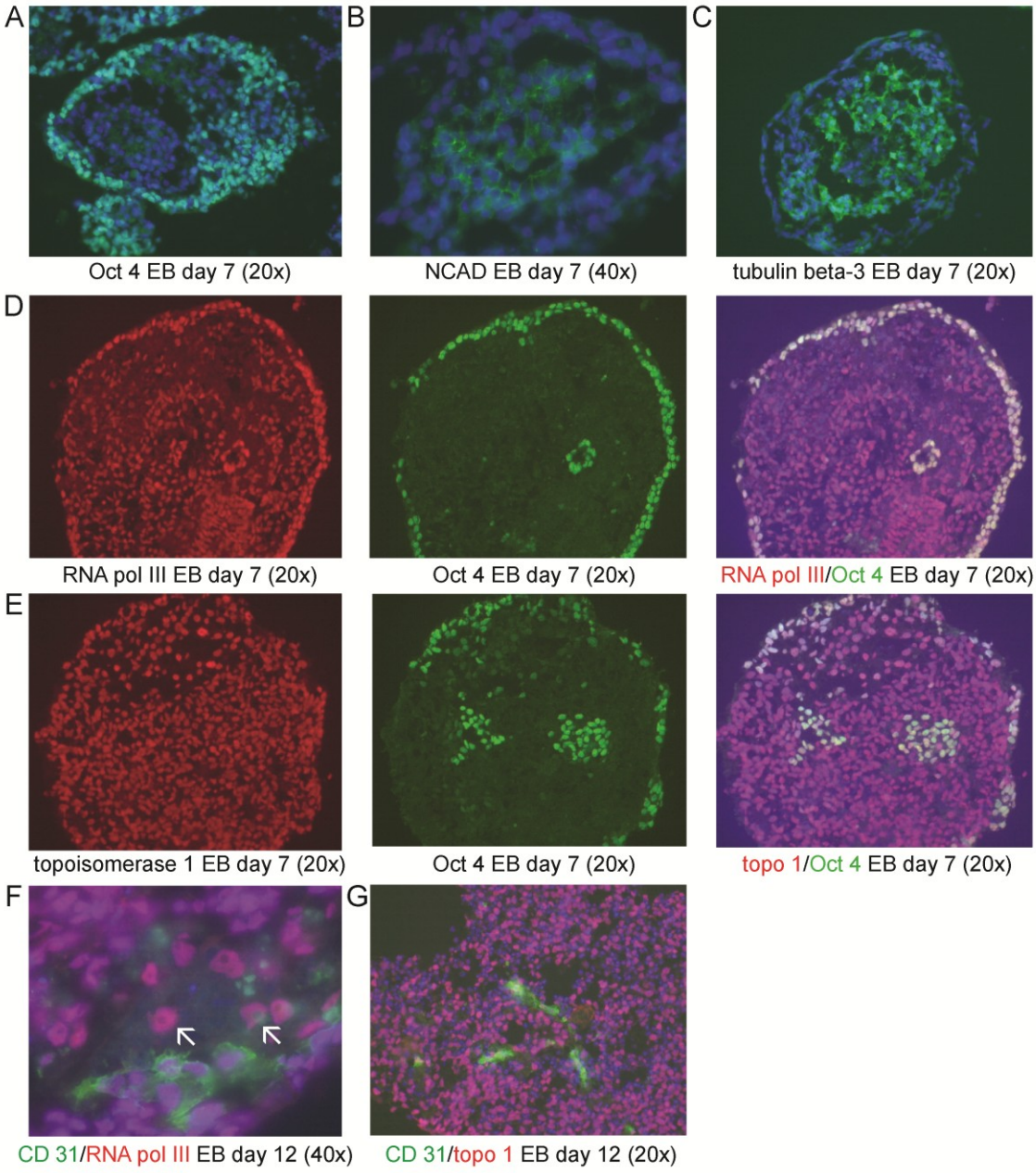

A
pH3----------//->pH10
116
97
66
45
B
pH3----------//->pH10
116
97
66
45
C
pH3----------//->pH10
116
97
66
45
D
HSC70
SSc-4
Overlay
E
SSc-4
Anti-fetuin A
Reduction + - - +
F
Deglycosylation
- +
Anti-fetuin A
SSc-4
G
Transfection
- +

Figure 14: Localization of lineage markers and scleroderma autoantigen expression in differentiating embryoid bodies. (A) Indirect immunofluorescence of Oct 4 expression in day seven EBs, with highest levels of expression along the periphery. (B) mesoendodermal marker NCAD and (C) ectodermal marker Tubulin beta-3 are expressed by cells located centrally in day seven EBs. (D) Colocalization of RNA polymerase III expression (red) with Oct 4 (green) demonstrates that RNA polymerase III is upregulated in cells expressing Oct 4 in day seven EBs. (E) Topoisomerase 1 expression is relatively uniform in day seven EBs and is not upregulated in Oct 4 expressing cells. (F & G) Co-localization of the endothelial cell marker CD31 (green) in day 12 EBs with RNA pol III (F, red) and topoisomerase 1 (G, red). Upregulation of RNA Pol III expression is detectable in cells of undetermined lineage (F, white arrows) located near the site of vessel formation. Cell nuclei are marked with DAPI (blue).


A
B
C
Oct 4 EB day 7 (20x)
NCAD EB day 7 (40x)
tubulin beta-3 EB day 7 (20x)
D
RNA pol III EB day 7 (20x)
Oct 4 EB day 7 (20x)
RNA pol III/Oct 4 EB day 7 (20x)
E
topoisomerase 1 EB day 7 (20x)
Oct 4 EB day 7 (20x)
topo 1/Oct 4 EB day 7 (20x)
F
G
CD 31/RNA pol III EB day 12 (40x)
CD 31/topo 1 EB day 12 (20x)

Chapter 4: General Conclusions

The systemic autoimmune diseases are a group of complex, heterogeneous diseases united by the presence of immune mediated destruction of target tissues and the presence of high titer, disease specific autoantibodies that correlate with clinical phenotype and prognosis. Evidence to date suggests that key insights into mechanisms of disease initiation and propagation can be obtained by characterizing the expression and modification of these targeted proteins in the cancers and target tissues of patients.

This investigation sought to further our understanding of these complex diseases by (1) the development of an improved approach for the identification of novel autoantigens that emphasizes the use of disease relevant antigen sources and a multidimensional approach that preserves and takes advantage of potentially relevant protein dynamics and (2) applying these methods to test the hypothesis that differentiating progenitor cells are targets of the immune response in scleroderma. This work demonstrated that current technologies enable rapid and efficient autoantigen discovery, and allow novel autoantigens expressed in pathways relevant to disease pathogenesis to be defined. Moreover, the detection of novel scleroderma autoantigens expressed in differentiating stem cells is the first evidence to suggest that a progenitor cell involved in tissue repair may be a target of the immune response in this disease.

Our current model of systemic autoimmunity is that an anti-tumor immune response cross-reacts with progenitors involved in tissue repair. In the setting of tissue damage, this cross reactivity results in a cycle of chronic antigen driven immune activation that disrupts the tissue repair process. This model explains features of these

diseases that have confounded researchers and clinicians alike, including: (*i*) The association with an increased risk of cancer, the diagnosis of which is temporally related to the diagnosis of autoimmunity, and the prognosis of which is better than that of the general population; (*ii*) The presence of autoantibodies years prior to the onset of autoimmune disease in some patients; (*iii*) The molecular and clinical heterogeneity of these diseases, likely reflecting the heterogeneity of the initiating tumor cells and anti-tumor lymphocytes; (*iv*) The self-sustaining and self-amplifying nature of the antigen driven immune response; (*v*) The chronic inflammation and tissue damage that subvert the body's homeostatic regulation and reparative efforts; and (*vi*) The confounding results of experiments identifying autoantigens using traditional cell lines. This model also accounts for the fact that despite the increased risk associated with systemic autoimmune diseases, a majority of patients are never diagnosed with cancer. Successful elimination of a nascent tumor and the development of a repertoire of autoreactive lymphocytes that go on to cause disease may be the intended and unintended consequences of an effective antitumor immune response.

Autoantibodies found in patient sera are a key link between anti-tumor and anti-self immune responses. They are also powerful probes for identifying the antigens responsible for driving the immune response, which may lead to the discovery of the cellular targets in systemic autoimmunity. Discovery of these targeted proteins and their role in the mechanism of disease offers hope for new clinical measures of disease as well as the development of desperately needed disease-specific therapies.

Chapter 5: References

1. A. Davidson, B. Diamond, in The Autoimmune Diseases., N. R. Rose, I. R. Mackay, Eds. (Elsevier Academic Press, St. Louis, 2006), pp. 25-11.

2. C. A. von Muhlen, E. M. Tan, Autoantibodies in the diagnosis of systemic rheumatic diseases. *Semin. Arthritis Rheum.* **24**, 323-358 (1995).

3. M. M. Nielen *et al.*, Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. *Arthritis Rheum.* **50**, 380-386 (2004).

4. M. R. Arbuckle *et al.*, Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N. Engl. J. Med.* **349**, 1526-1533 (2003).

5. T. Kawasaki, T. Kawai, S. Akira, Recognition of nucleic acids by pattern-recognition receptors and its relevance in autoimmunity. *Immunol. Rev.* **243**, 61-73 (2011).

6. E. Rai, E. K. Wakeland, Genetic predisposition to autoimmunity--what have we learned? *Semin. Immunol.* **23**, 67-83 (2011).

7. E. E. Sercarz *et al.*, Dominance and crypticity of T cell antigenic determinants. *Annu. Rev. Immunol.* **11**, 729-766 (1993).

8. A. Lanzavecchia, How can cryptic epitopes trigger autoimmunity? *J. Exp. Med.* **181**, 1945-1948 (1995).

9. H. A. Doyle, M. J. Mamula, Posttranslational modifications of self-antigens. *Ann. N. Y. Acad. Sci.* **1050**, 1-9 (2005).

10. P. Eggleton, R. Haigh, P. G. Winyard, Consequence of neo-antigenicity of the 'altered self'. *Rheumatology (Oxford).* **47**, 567-571 (2008).

11. E. Darrah, A. Rosen, Granzyme B cleavage of autoantigens in autoimmunity. *Cell Death Differ.* **17**, 624-632 (2010).

12. M. E. Engelhorn *et al.*, Autoimmunity and tumor immunity induced by immune responses to mutations in self. *Nat. Med.* **12**, 198-206 (2006).

13. L. Casciola-Rosen *et al.*, Enhanced autoantigen expression in regenerating muscle cells in idiopathic inflammatory myopathy. *J. Exp. Med.* **201**, 591-601 (2005).

14. M. Mahler, F. W. Miller, M. J. Fritzler, Idiopathic inflammatory myopathies and the anti-synthetase syndrome: A comprehensive review. *Autoimmun. Rev.*(2014).

15. A. L. Mammen *et al.*, Autoantibodies against 3-hydroxy-3-methylglutaryl-coenzyme A reductase in patients with statin-associated autoimmune myopathy. *Arthritis Rheum.* **63**, 713-721 (2011).

16. C. L. Hill *et al.*, Frequency of specific cancer types in dermatomyositis and polymyositis: a population-based study. *Lancet.* **357**, 96-100 (2001).

17. R. Buchbinder, A. Forbes, S. Hall, X. Dennett, G. Giles, Incidence of malignant disease in biopsy-proven inflammatory myopathy. A population-based cohort study. *Ann. Intern. Med.* **134**, 1087-1095 (2001).

18. S. Bernatsky *et al.*, The relationship between cancer and medication exposures in systemic lupus erythaematosus: a case-cohort study. *Ann. Rheum. Dis.* **67**, 74-79 (2008).

19. S. Bernatsky *et al.*, An international cohort study of cancer in systemic lupus erythematosus. *Arthritis Rheum.* **52**, 1481-1490 (2005).

20. M. N. Lazarus, D. Robinson, V. Mak, H. Moller, D. A. Isenberg, Incidence of cancer in a cohort of patients with primary Sjogren's syndrome. *Rheumatology (Oxford).* **45**, 1012-1015 (2006).

21. F. Wolfe, K. Michaud, Lymphoma in rheumatoid arthritis: the effect of methotrexate and anti-tumor necrosis factor therapy in 18,572 patients. *Arthritis Rheum.* **50**, 1740-1751 (2004).

22. F. Wolfe, K. Michaud, Biologic treatment of rheumatoid arthritis and the risk of malignancy: analyses from a large US observational study. *Arthritis Rheum.* **56**, 2886-2895 (2007).

23. M. Wooten, Systemic sclerosis and malignancy: a review of the literature. *South. Med. J.* **101**, 59-62 (2008).

24. M. Colaci *et al.*, Lung cancer in scleroderma: results from an Italian rheumatologic center and review of the literature. *Autoimmun. Rev.* **12**, 374-379 (2013).

25. E. Toubi, Y. Shoenfeld, Protective autoimmunity in cancer (review). *Oncol. Rep.* **17**, 245-251 (2007).

26. G. Q. Phan *et al.*, Cancer regression and autoimmunity induced by cytotoxic T lymphocyte-associated antigen 4 blockade in patients with metastatic melanoma. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 8372-8377 (2003).

27. H. Gogas *et al.*, Prognostic significance of autoimmunity during treatment of melanoma with interferon. *N. Engl. J. Med.* **354**, 709-718 (2006).

28. L. M. Stinton, M. J. Fritzler, A clinical approach to autoantibody testing in systemic autoimmune rheumatic disorders. *Autoimmun. Rev.* **7**, 77-84 (2007).

29. N. Bizzaro, Autoantibodies as predictors of disease: the clinical and experimental evidence. *Autoimmun. Rev.* **6**, 325-333 (2007).

30. A. L. Mammen, Autoimmune myopathies: autoantibodies, phenotypes and pathogenesis. *Nat. Rev. Neurol.* **7**, 343-354 (2011).

31. L. Casciola-Rosen, F. Andrade, D. Ulanet, W. B. Wong, A. Rosen, Cleavage by granzyme B is strongly predictive of autoantigen status: implications for initiation of autoimmunity. *J. Exp. Med.* **190**, 815-826 (1999).

32. A. L. Mammen *et al.*, Expression of the dermatomyositis autoantigen Mi-2 in regenerating muscle. *Arthritis Rheum.* **60**, 3784-3793 (2009).

33. D. Fiorentino, L. Chung, J. Zwerner, A. Rosen, L. Casciola-Rosen, The mucocutaneous and systemic phenotype of dermatomyositis patients with antibodies to MDA5 (CADM-140): a retrospective study. *J. Am. Acad. Dermatol.* **65**, 25-34 (2011).

34. S. Sato *et al.*, RNA helicase encoded by melanoma differentiation-associated gene 5 is a major autoantigen in patients with clinically amyopathic dermatomyositis: Association with rapidly progressive interstitial lung disease. *Arthritis Rheum.* **60**, 2193-2200 (2009).

35. S. A. Greenberg *et al.*, Relationship between disease activity and type 1 interferon- and other cytokine-inducible gene expression in blood in dermatomyositis and polymyositis. *Genes Immun.*(2011).

36. L. Canelle *et al.*, An efficient proteomics-based approach for the screening of autoantibodies. *J. Immunol. Methods.* **299**, 77-89 (2005).

37. P. Mallick, B. Kuster, Proteomics: a pragmatic perspective. *Nat. Biotechnol.* **28**, 695-709 (2010).

38. M. W. Duncan, R. Aebersold, R. M. Caprioli, The pros and cons of peptide-centric proteomics. *Nat. Biotechnol.* **28**, 659-664 (2010).

39. A. Bohan, J. B. Peter, Polymyositis and dermatomyositis (second of two parts). *N. Engl. J. Med.* **292**, 403-407 (1975).

40. A. Bohan, J. B. Peter, Polymyositis and dermatomyositis (first of two parts). *N. Engl. J. Med.* **292**, 344-347 (1975).

41. W. Pan, A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics.* **18**, 546-554 (2002).

42. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological).* **57**, 289 (1995).

43. S. A. Greenberg *et al.*, Interferon-alpha/beta-mediated innate immune mechanisms in dermatomyositis. *Ann. Neurol.* **57**, 664-678 (2005).

44. R. J. Prescott, A. J. Freemont, C. J. Jones, J. Hoyland, P. Fielding, Sequential dermal microvascular and perivascular changes in the development of scleroderma. *J. Pathol.* **166**, 255-263 (1992).

45. D. J. Abraham, T. Krieg, J. Distler, O. Distler, Overview of pathogenesis of systemic sclerosis. *Rheumatology (Oxford).* **48 Suppl 3**, iii3-7 (2009).

46. O. Distler *et al.*, Uncontrolled expression of vascular endothelial growth factor and its receptors leads to insufficient skin angiogenesis in patients with systemic sclerosis. *Circ. Res.* **95**, 109-116 (2004).

47. J. H. Distler, S. Gay, O. Distler, Angiogenesis and vasculogenesis in systemic sclerosis. *Rheumatology (Oxford).* **45 Suppl 3**, iii26-7 (2006).

48. A. Gabrielli, E. V. Avvedimento, T. Krieg, Scleroderma. *N. Engl. J. Med.* **360**, 1989-2003 (2009).

49. Y. S. Gu *et al.*, The immunobiology of systemic sclerosis. *Semin. Arthritis Rheum.* **38**, 132-160 (2008).

50. X. Zhou *et al.*, HLA-DPB1 and DPB2 are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. *Arthritis Rheum.* **60**, 3807-3814 (2009).

51. F. C. Arnett *et al.*, Major histocompatibility complex (MHC) class II alleles, haplotypes and epitopes which confer susceptibility or protection in systemic sclerosis: analyses in 1300 Caucasian, African-American and Hispanic cases and 1000 controls. *Ann. Rheum. Dis.* **69**, 822-827 (2010).

52. B. W. Needleman, F. M. Wigley, R. W. Stair, Interleukin-1, interleukin-2, interleukin-4, interleukin-6, tumor necrosis factor alpha, and interferon-gamma levels in sera from patients with scleroderma. *Arthritis Rheum.* **35**, 67-72 (1992).

53. L. I. Sakkas *et al.*, Oligoclonal T cell expansion in the skin of patients with systemic sclerosis. *J. Immunol.* **168**, 3649-3659 (2002).

54. V. D. Steen, Autoantibodies in systemic sclerosis. *Semin. Arthritis Rheum.* **35**, 35-42 (2005).

55. E. J. Cepeda, J. D. Reveille, Autoantibodies in systemic sclerosis and fibrosing syndromes: clinical indications and relevance. *Curr. Opin. Rheumatol.* **16**, 723-732 (2004).

56. A. A. Shah, A. Rosen, L. Hummers, F. Wigley, L. Casciola-Rosen, Close temporal relationship between onset of cancer and scleroderma in patients with RNA polymerase I/III antibodies. *Arthritis Rheum.* **62**, 2787-2795 (2010).

57. C. G. Joseph *et al.*, Association of the autoimmune disease scleroderma with an immunologic response to cancer. *Science.* **343**, 152-157 (2014).

58. D. J. Laird, U. H. von Andrian, A. J. Wagers, Stem cell trafficking in tissue development, growth, and disease. *Cell.* **132**, 612-630 (2008).

59. D. J. Wong *et al.*, Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell. Stem Cell.* **2**, 333-344 (2008).

60. I. Ben-Porath *et al.*, An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet.* **40**, 499-507 (2008).

61. M. A. Nieto, Epithelial plasticity: a common theme in embryonic and cancer cells. *Science.* **342**, 1234850 (2013).

62. M. Mimeault, S. K. Batra, Concise review: recent advances on the significance of stem cells in tissue regeneration and cancer therapies. *Stem Cells.* **24**, 2319-2345 (2006).

63. H. Kurosawa, Methods for inducing embryoid body formation: in vitro differentiation system of embryonic stem cells. *J. Biosci. Bioeng.* **103**, 389-398 (2007).

64. T. Dvash, N. Benvenisty, Human embryonic stem cells as a model for early human development. *Best Pract. Res. Clin. Obstet. Gynaecol.* **18**, 929-940 (2004).

65. O. Kopper, O. Giladi, T. Golan-Lev, N. Benvenisty, Characterization of gastrulation-stage progenitor cells and their inhibitory crosstalk in human embryoid bodies. *Stem Cells.* **28**, 75-83 (2010).

66. A. T. Masi *et al.*, Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. *Arthritis Rheum.* **23**, 581-590 (1980).

67. Q. Xi *et al.*, A poised chromatin platform for TGF-beta access to master regulators. *Cell.* **147**, 1511-1524 (2011).

68. J. Li, H. Stuhlmann, In vitro imaging of angiogenesis using embryonic stem cell-derived endothelial cells. *Stem Cells Dev.* **21**, 331-342 (2012).

69. K. Hodge, S. T. Have, L. Hutton, A. I. Lamond, Cleaning up the masses: exclusion lists to reduce contamination with HPLC-MS/MS. *J. Proteomics.* **88**, 92-103 (2013).

70. D. K. Trask *et al.*, Keratins as markers that distinguish normal and tumor-derived mammary epithelial cells. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 2319-2323 (1990).

71. E. Mizuuchi, S. Semba, Y. Kodama, H. Yokozaki, Down-modulation of keratin 8 phosphorylation levels by PRL-3 contributes to colorectal carcinoma progression. *Int. J. Cancer.* **124**, 1802-1810 (2009).

72. T. A. Hembrough, L. Li, S. L. Gonias, Cell-surface cytokeratin 8 is the major plasminogen receptor on breast cancer cells and is required for the accelerated activation of cell-associated plasminogen by tissue-type plasminogen activator. *J. Biol. Chem.* **271**, 25684-25691 (1996).

73. M. C. Moon, L. Yau, B. Wright, P. Zahradka, Injury-induced expression of cytokeratins 8 and 18 by vascular smooth muscle cells requires concurrent activation of cytoskeletal and growth factor receptors. *Can. J. Physiol. Pharmacol.* **86**, 223-231 (2008).

74. N. Marceau, A. Loranger, S. Gilbert, N. Daigle, S. Champetier, Keratin-mediated resistance to stress and apoptosis in simple epithelial cells in relation to health and disease. *Biochem. Cell Biol.* **79**, 543-555 (2001).

75. R. G. Oshima, Apoptosis and keratin intermediate filaments. *Cell Death Differ.* **9**, 486-492 (2002).

76. C. Choudhary *et al.*, Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science.* **325**, 834-840 (2009).

77. S. H. Leech *et al.*, Proteomic analyses of intermediate filaments reveals cytokeratin8 is highly acetylated--implications for colorectal epithelial homeostasis. *Proteomics.* **8**, 279-288 (2008).

78. C. F. Chou, A. J. Smith, M. B. Omary, Characterization and dynamics of O-linked glycosylation of human cytokeratin 8 and 18. *J. Biol. Chem.* **267**, 3901-3906 (1992).

79. J. P. Lebreton *et al.*, Serum concentration of human alpha 2 HS glycoprotein during the inflammatory process: evidence that alpha 2 HS glycoprotein is a negative acute-phase reactant. *J. Clin. Invest.* **64**, 1118-1129 (1979).

80. A. M. Sakwe, R. Koumangoye, S. J. Goodwin, J. Ochieng, Fetuin-A ({alpha}2HS-glycoprotein) is a major serum adhesive protein that mediates growth signaling in breast tumor cells. *J. Biol. Chem.* **285**, 41827-41835 (2010).

81. Z. Nie, Fetuin: its enigmatic property of growth promotion. *Am. J. Physiol.* **263**, C551-62 (1992).

82. D. Caballero-Hernandez, R. Gomez-Flores, P. Tamez-Guerra, R. Tamez-Guerra, C. Rodriguez-Padilla, Role of immunogenic fetuin A on L5178Y-R lymphoma tumorigenesis. *Cancer Invest.* **27**, 257-263 (2009).

83. B. Guillory *et al.*, Lack of fetuin-A (alpha2-HS-glycoprotein) reduces mammary tumor incidence and prolongs tumor latency via the transforming growth factor-beta signaling pathway in a mouse model of breast cancer. *Am. J. Pathol.* **177**, 2635-2644 (2010).

84. J. K. Yi *et al.*, Autoantibody to tumor antigen, alpha 2-HS glycoprotein: a novel biomarker of breast cancer screening and diagnosis. *Cancer Epidemiol. Biomarkers Prev.* **18**, 1357-1364 (2009).

85. D. Launay *et al.*, Association between systemic sclerosis and breast cancer: eight new cases and review of the literature. *Clin. Rheumatol.* **23**, 516-522 (2004).

86. C. Schafer *et al.*, The serum protein alpha 2-Heremans-Schmid glycoprotein/fetuin-A is a systemically acting inhibitor of ectopic calcification. *J. Clin. Invest.* **112**, 357-366 (2003).

87. A. Heiss *et al.*, Structural basis of calcification inhibition by alpha 2-HS glycoprotein/fetuin-A. Formation of colloidal calciprotein particles. *J. Biol. Chem.* **278**, 13333-13341 (2003).

88. Chapter 1 was adapted with permission from: T. Cottrell and A. Rosen, Mechanisms of Autoimmunity. In Clinical Immunology: Principles and Practice (4th edition), R. Rich, et al., Eds. Ch 48, pp 587-94. Copyright Elsevier (2013).

89. Chapter 2 was previously published in *J. Immunol. Methods*. **385**, 35-44. T. R. Cottrell, J. C. Hall, A. Rosen, L. Casciola-Rosen, Identification of novel autoantigens by a triangulation approach. Copyright Elsevier (2012).

Tricia Ruth Cottrell
4619 Warren Tree Way
Baltimore, MD 21229
tcottre2@jhmi.edu
410-218-0773

Birth

St. Louis, Missouri 12/10/1981

Education

MD/PhD Candidate, May 2014
Johns Hopkins University School of Medicine
Graduate Program in Cellular and Molecular Medicine
Baltimore, MD 2004-Present

Bachelor of Arts, May 2004
Washington University in St. Louis, School of Arts and Sciences
Chemistry (Biochemistry Concentration) & English Literature
St. Louis, MO 2000-2004

Study Abroad Program
University College London, Department of English Literature
London, United Kingdom 1/2003-6/2003

Teaching Experience

High School Math and Science Tutor
Friends School of Baltimore, Baltimore, MD 2008-2010

Cell Physiology Teaching Assistant
Johns Hopkins School of Medicine, Baltimore, MD 2006-2008

Histology Teaching Assistant
Johns Hopkins School of Medicine, Baltimore, MD 2007

Research Experience

PhD Candidate, PI: Dr. Antony Rosen
Johns Hopkins School of Medicine, Baltimore, MD 2006-2012

Undergraduate Research Fellow, PI: Dr. Tamara Doering

Washington University School of Medicine, St. Louis, MO 2000-2004

Honors

ACR REF Medical/Graduate Student Achievement Award	2010
Phi Beta Kappa, Washington University in St. Louis	2004
Sigma Xi, Washington University in St. Louis	2004

Funding

Arnold and Mabel Beckman Research Scholarship	2003-2004
American Society for Microbiology Summer Research Fellowship	2001, 2002
Howard Hughes Medical Institute Summer Research Fellowship	2001
HHMI Prefreshman Biology and Biomedical Research Program	2000

Publications

Cottrell TR, Wise RA, Wigley FM, Boin F. The degree of skin involvement identifies distinct lung disease outcomes and survival in systemic sclerosis. Ann Rheum Dis. 2013 Apr 20. [Epub ahead of print]

Cottrell TR, Rosen A. Mechanisms of autoimmunity. In: Rich R, Fleisher T, William S, Schroeder H, Frew A, Weyand C, editors. Clinical Immunology: Principles and Practice (4th edition); 2013. p. 587-594.

Cottrell TR, Hall JC, Rosen A, Casciola-Rosen L. Identification of novel autoantigens by a triangulation approach. J Immunol Methods. 2012 Nov 30;385(1-2):35-44

Cottrell TR, Griffith CL, Liu H, Nenninger AA, Doering TL. "The pathogenic fungus Cryptococcus neoformans expresses two functional GDP-mannose transporters with distinct expression patterns and roles in capsule synthesis." Eukaryotic Cell. 2007 May;6(5):776-85.

Cottrell TR and Doering TL. "Silence of the Strands: RNA Interference in Eukaryotic Pathogens." Trends in Microbiology. 2003 Jan;11(1):37-43.

Liu H, Cottrell TR, Perini LM, Goldman WE, and Doering TL. "RNA Interference in the Pathogenic Fungus Cryptococcus neoformans." Genetics. 2002 Feb;160(2):463-70.

Poster Presentations

Cottrell, T., Wigley, F., Wise, R., Boin, F. (2010). The Degree of Skin Involvement Predicts Distinct Interstitial Lung Disease Outcomes in Systemic Sclerosis. Poster

presented at: American College of Rheumatology Annual Scientific Meeting; Atlanta, GA.

Cottrell, T., Liu, H., Bose, I., Pierini, L., Goldman, W., Doering, T. (2002). RNA Interference in Cryptococcus neoformans. Poster presented at: American Society for Microbiology General Meeting; Salt Lake City, UT.

Cottrell, T., Liu, H., Bose, I., Pierini, L., Goldman, W., Doering, T. (2002). RNA Interference in Cryptococcus neoformans. Poster presented at: Midwest Microbial Pathogenesis Conference; Indianapolis, IN.

Oral Presentations

Cottrell, T., Rosen, L., Rosen, A. (2009). Scleroderma Autoantibodies: New Insight. Oral Presentation presented at: Scleroderma Advisory Board; Baltimore, MD.

Cottrell, T., Liu, H., Bose, I., Pierini, L., Goldman, W., Doering, T. (2002). Use of dsRNA interference to suppress gene expression in Cryptococcus neoformans. Oral Presentation presented at: WUSM Infectious Disease/Basic Microbiological Mechanisms Research Conference; St. Louis, MO.

Cottrell, T., Liu, H., Pierini, L., Goldman, W., Doering, T. (2001). New Methods for Modulation and Quantitation of Gene Expression in Cryptococcus neoformans. Oral Presentation presented at: HHMI Summer Research Program Conference; St. Louis, MO.

Cottrell, T., Liu, H., Pierini, L., Goldman, W., Doering, T. (2001). Use of dsRNA interference to suppress gene expression in Cryptococcus neoformans. Oral Presentation presented at: WUSM Medical Mycology Research Club; St. Louis, MO.

Cottrell, T., Liu, H., Doering, T. (2000). RNA Interference in Cryptococcus neoformans. Oral Presentation presented at: HHMI Prefreshman Biology and Biomedical Research Conference; St. Louis, MO.